二十世纪西方哲学译丛

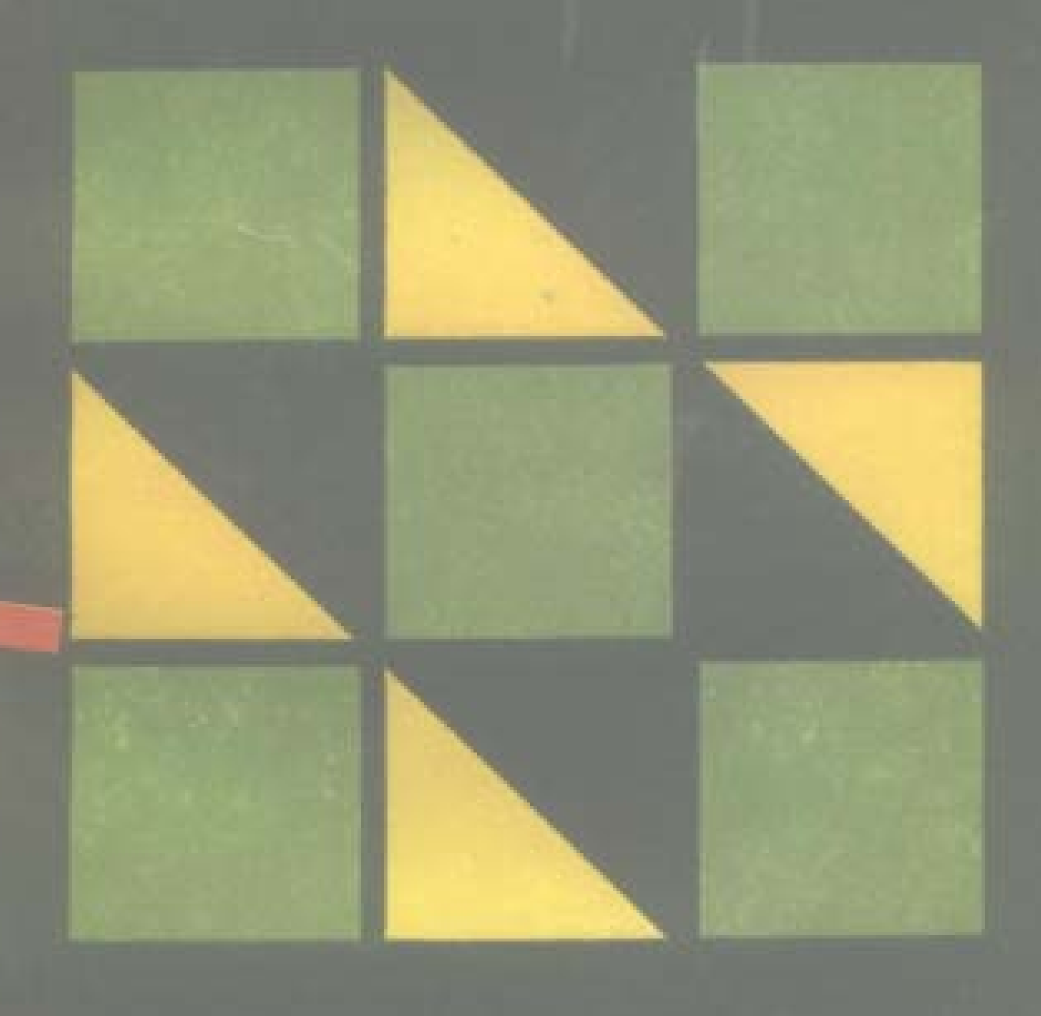
心、脑与科学

Minds, Brains and Science

[美] 约翰·塞尔 著

杨音莱 译

上海译文出版社



二十世纪西方哲学译丛

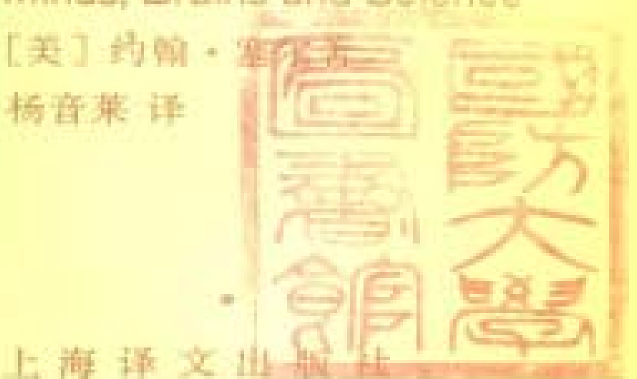




心脑与科学

Minds, Brains and Science

[美] 约翰 · 塞尔
杨音莱 译



上海译文出版社



John Searle

MINDS, BRAINS AND SCIENCE

Harvard University Press, 1984

根据哈佛大学出版社1984年版译出

心、脑与科学

〔美〕约翰·塞尔 著

杨音莱 译

上海译文出版社出版、发行

上海延安中路955弄14号

全国新华书店经销

上海译文印刷厂印刷

开本850×1168 1/32 印张[illegible] 插页[illegible] 字数[illegible]

1991年10月第1版 1991年10月第1次印刷

印数：[illegible]册

ISBN7-5327-1153-6/B·064

定价：3.05元

（沪）新登字111号

中译本作者序言

我高兴地看到《心、脑与科学》一书中译本出版。中国是具有古老和优秀哲学传统的国度，今天她正涌动着巨大的理智活力。也许，本书的出版能够对中西方哲学家之间与日俱增的交流作出一点儿小小的贡献。

John Searle

译者的话

美国当代著名的分析哲学家约翰·罗杰斯·塞尔1932年生于美国丹佛市，1949年至1952年就读于美国威斯康星大学，1955年赴英国牛津大学留学。在牛津，他一边在著名的日常语言哲学家约翰·奥斯汀的指导下研究语言分析哲学，一边在牛津基督教堂学院任教。1959年，获文学硕士和哲学博士学位，并于同年返美。以后，他一直在美国加州大学伯克利分校任教，曾担任过哲学系主任、校长特别助理等职。现在，他是该校负有盛名的哲学教授，美国科学院的院士。

塞尔首先是一位以研究语言哲学问题著称的分析哲学家。多年来他一直在潜心研究由他的老师约翰·奥斯汀提出的言语行为理论。在这方面，他的研究专著有《言语行为：语言哲学方面的一篇论文》(1969年)、《词语与意义：言语行为理论研究》(1979年)和与丹尼尔·范德维肯合著的《非表现逻辑基础》(1985年)。此外，他还主编了《语言哲学》(1971年)和《言语行为理论和语用学》两本论文集，并撰写了大量有关论文。他的这些研究成果，使他成为当今英美哲学界继约翰·奥斯汀以来该研究领域里公认的权威。

随着对言语行为理论研究的深入，塞尔越来越重视对这种理论背后更基本的哲学问题的研究。他认为，言语与人类的其他行为一样是一种行为，对语言的哲学分析就是对人的行为进行一般分析的一部分，而在言语行为以及人类其他类型的行为

背后更基本的东西是人心中的意向性。1983年出版了他的《意向性：心的哲学方面的一篇论文》一书，该书把意向性问题作为当代分析哲学的一个基本问题加以探讨。对意向性问题的研究又导致他从更一般的意义上去探讨人心的特性问题。在他看来，既然意向性行为是心理现象的一种表现，那么就表明语言哲学和行动哲学只不过是心的哲学这个更大领域中的一些不同的方面。心的哲学在当代英美分析哲学的各领域中是一个特别受到重视的领域。这不仅因为越来越多的分析哲学家认为曾被传统分析哲学视为哲学首要问题的语言哲学问题可以归结为心的哲学问题，而且还因为由于当代科学的发展引起的一系列新问题，像计算机能否思维、脑功能与意识现象的关系等问题也有待于心的哲学的回答。近十几年来，塞尔在心的哲学及其有关领域内发表了一系列论战性的文章，这使他从一个分析哲学某一领域里的专家成为一个涉足广泛、对许多哲学问题都比较敏感和富有挑战精神的分析哲学家。

本书比较集中地反映了塞尔对心的哲学问题以及与此有密切关系的几个哲学问题的研究成果和哲学见解。它的初稿是塞尔在1984年英国广播公司举办的里思学术系列广播讲座上的演讲稿，经修改后，以《心、脑与科学》为书名于当年出版。本书在内容、篇幅和风格上基本上保持了演讲稿的原样。书中六章所探讨的六个问题始终围绕着一个主题展开，这个主题就是从哲学上去探讨人类与周围宇宙的关系问题。塞尔认为这个问题是一个被今天的许多哲学家有意回避的“大问题”，也是当代哲学所面临的最大难题。在他看来，回答人类如何与周围宇宙统一的问题实际上就是一个如何把我们关于我们自己作为有意识的行为者的常识概念与我们把世界视作一个物理体系的科学概念统一起来的问题。他坚持认为，这种常识的概念和科学的概念

不仅彼此是不矛盾的，而且是完全一致的。书中所探讨的六个问题实际上就是从六个方面去说明和论证他的这一基本哲学立场的。前三章主要围绕着心的哲学的中心问题即心身问题展开，这个传统哲学难题在当代通常采取心脑关系问题的形式。在第一章，他从正面提出了解决心身问题的方法，并简明地表述了他称之为“朴素物理主义”和“朴素心理主义”的两个观点，坚持这两个观点的一致性是塞尔心的哲学立场的一个基本特征。在第二、三章，塞尔从他的这种基本观点和立场出发，驳斥了把计算机程序和硬件的关系等同于心脑关系的所谓“强人工智能”观点和试图在心脑之间插入一个信息加工层次的认知主义研究纲领。后三章，塞尔探讨了三个与心的哲学密切相关的重要问题。所谓“行动哲学”是分析哲学中一个兴盛的分支，其研究方向主要是从人的心理特性及其与身体特性之间的关系角度来分析人的行动结构。塞尔把人心指向外界事物的意向性视为贯穿在人的行动中最基本、最重要的特性，以意向性概念为核心展示了一个具有一般意义的行动结构。社会科学的哲学在当前英美哲学界是一个争论颇多的领域，塞尔在把他的心的哲学和行动哲学观点应用到社会哲学问题上提出了他的独到见解。他认为社会科学的概念具有一种“自我涉及”的性质，因此在自然科学概念与社会科学概念之间存在着一种“根本的中断”；社会科学“涉及意向性的种种方面”，它研究的是人们在社会生活中的种种意向性。最后，他向自由意志问题这个“典型的哲学之谜”提出挑战。他以一种“修正的相容论”观点试图协调决定论与自由意志论之间的冲突，认为我们关于物理世界的科学概念与人们对自由意志的确信这两方面都有着不可驳倒的事实，并主张“物理的决定论”与“心理的自由意志论”是能够彼此相容的。

毫无疑问，本书是作者阐述自己哲学世界观的一部重要著

作。国外哲学评论家认为这本小册子比作者以往任何著作都更广、更深地表述了他的心的哲学观点。① 本书出版以来，受到英美哲学界的普遍重视。这首先是因为作者在书中提出了一种试图把常识与科学统一起来的哲学世界观。这种世界观不同于传统分析哲学一贯主张的还原论的哲学立场，也不同于传统二元论以及当代行为主义、功能主义的哲学立场。尽管塞尔的这种哲学立场并没有得到分析哲学家们的一致赞同，但它毕竟是当代分析哲学对待哲学基本问题所持有的一种态度，它表明了分析哲学在不断发展中的一种新动向。其次，塞尔在解决每个具体问题的过程中，使用了一些具有独创性的论证方法。例如，在驳斥"强人工智能"观点时，为了论证数字计算机永远不可能具有心灵这一论点，他巧妙地举了"汉语屋子"的例子。这个例子已成为当代西方哲学家们在讨论有关问题时经常引用的一个经典例证。此外，在解决心身问题时，他提出了一种新的因果关系理论，以此来取代那种只能描述外在、分离事件的机械的因果模式；在评价认知主义理论纲领时，他使用了将"遵守规则"与"信息加工"两个概念相类比的论证方法，等等。这些论证方法对丰富分析哲学的理论方法有着不可忽视的学术价值。当然，对本书所阐述的哲学立场和观点有待于人们进一步作出评价。应该指出的是，站在塞尔所坚持的这种哲学立场上是否能够正确地解决所谓人类与周围宇宙关系的"大难题"，这个根本问题值得我们进一步思考。本书也是作者力图通俗地介绍他近年来研究成果的一次尝试。作者对所涉及到的各种不同学术观点尽量加以简明、通俗地介绍，在与不同观点的论战中提出他自己的

① 参阅《哲学评论》(Philosophical Review)，1984年，第1期，第129页。《元哲学》(Metaphilosophy)，1986年，第2期，第172页。

哲学观点，作者采取这种叙述方法有助于我们更好地了解当代英美哲学在这些领域里的理论现状和发展趋势。

本书中译本附录收入了塞尔的两篇重要论文，一篇讨论言语行为问题，另一篇讨论意向性问题。两篇论文比较典型地代表了塞尔对当代西方语言哲学和心的哲学有所贡献的两个方面，从这两个方面读者能更全面地了解塞尔的哲学思想。

1987年夏，当时在中国人民大学哲学系执教的江怡老师向我推荐了这本小书，并建议我将它译成中文；在翻译过程中，莹珊给予我莫大的支持和鼓励，在此一并致谢。

译文不妥之处，恳请读者指正。

杨音莱

1988年8月于北京沙滩

目　录

导 言

我十分荣幸地应邀作了1984年的里思讲演。继伯特兰·罗素在1948年开始了这个系列讲座之后，这还是第一次由一位哲学家来主讲。

但是，如果说作这个讲演是一种荣誉的话，那么它也是一种挑战。理想的里思系列讲座应包括六个广播单元，每个单元正好占半小时，各单元应能独立成篇，同时六个单元又能共同构成一个统一的整体。这个系列讲座应以先前主讲人的工作成果为基础，同时也应包含新的和具有独创性的内容。也许最难达到的就是，让那些有兴趣和敏感的听众能够完全听懂这个讲座，因为他们当中的多数人对论题内容、术语以及专家们持有的成见所知甚微。我不知道所有这些目标能否同时达到，但不管怎么说，这些正是我要争取做到的。促使我作里思讲演最主要的原因之一就是，我深信，现代分析哲学的成果与方法能为更广泛的听众所接受。

将讲演稿编写成书，我最初的打算是对各章作某些扩充，以回答所有我所能设想到的来自我的那些善辩的同行们的异议，更不必说来自我在认知科学、人工智能和其他领域的同事们的异议。简单地说，我原想把讲演稿编写成一部有注脚等部分的正规的著作。最后，我还是改变了这种想法，因为采取系列讲座的方式完全能为任何一个有很大兴趣想理解这些论证的人所接受，而如果像原先计划的那样改写，将首先破坏了这个在我看来

是本系列讲座中最富有感染力的一个方面。因此，书中的各章基本上是我里思讲演稿的原样。只是为了让内容更清楚明白，我作了些扩充，但我力图保持当初讲稿的文体、语调和那种不拘形式的风格。

这个系列讲座中压倒一切的主题涉及到人类与周围宇宙的关系，它尤其关系到我们如何去使以下两种概念相一致的问题：一种是关于我们自身的传统的心理主义的概念；一种是视宇宙为一个纯物理系统或一组相互作用的物理系统的表面看来不一致的概念。围绕这一主题，每一章讲一个具体的问题：什么是心与脑的关系？单凭有合适输入与输出的恰当程序，数字计算机能否具有心灵？一种计算机程序的心理模式是否可能？什么是人类行为结构的性质？社会科学作为科学，它的地位如何？如果完全可能的话，我们怎样将我们对自己自由意志的确信与我们视宇宙为一个或相互作用的一组物理系统的观念协调一致？

在整个系列演讲过程中，还涉及到某些其他重要问题，不过由于篇幅所限，不能充分展开论述。因此我想在这个导言中把这些问题充分地说明一下，这样做能够帮助读者更好地理解以下各章的内容。

第一点我要说明的是，关于人脑的功能我们所知甚少，而基于这种无知之上的某些理论的矜夸与造作又是如此之多。神经生理学家戴维·休伯尔在1978年写道："我们关于脑的知识处于一种极其初级的程度。尽管对一些领域我们已提出了某种功能的概念，然而其他方面，就掌握的程度而言，我们的认识几乎相当于尚未知道心脏有泵血功能时对心脏的认识水平。"的确，如果一个对此感兴趣的外行人，像我这样，读了半打关于脑的权威教科书，并且努力去寻求能使任何好奇的人立即想到的那些问题的答案，那么，他很可能会感到失望的。究竟什么是意识的神

经生理学？为什么我们需要睡眠？究竟为什么酒精会使我们变醉？记忆究竟是怎样贮存在大脑中的？到目前为止，我们简直不知道任何关于这些基本问题的答案。从弗洛伊德的心理学到人工智能等种种学科提出的许多关于心灵的主张正是建立在这种无知之上的。这些主张是建立在我们极贫乏的知识基础之上的。

将神经原作为脑功能的基本单位，是对脑的传统看法，这不过是有关脑功能最显著的东西。脑接收的所有输入种类，包括刺激视网膜的光、刺激耳膜的声波、作用于皮肤而使神经末梢活动引起触压、热、冷和痛觉的压力等等，所有这些信号采用一种共同的输入手段：各种不同速率的神经原冲动。同样显而易见的是，这些不同速率的神经原冲动，在各神经环路中和脑的各部位状态下引起所有那些形形色色的心理生活。一朵玫瑰花的气味、对蓝色天空的体察、洋葱的味道以及数学公式的观念，这一切都产生于各环路中、与脑的各部位状态相关的不同速率的神经原冲动。

那么造成我们心理生活差异的种种神经环路和局部环境究竟是什么呢？没有人能详细地回答这个问题，但我们确实有可靠的证据说明脑的一定区域被一定种类的经验所分化。视觉皮质在视觉经验中有着一种专门的功能，听觉皮质在听觉经验中也是如此，等等。假如将听觉刺激提供给视觉皮质，将视觉刺激提供给听觉皮质，那么将会产生什么结果呢？据我所知，没有人做过这种实验。但似乎这样的假设是有道理的：听觉刺激会被“看见”，也就是说，听觉刺激会引起视觉经验；视觉刺激会被“听见”，即，视觉刺激会引起听觉经验；这两种现象来自视、听觉皮质各自的、我们还不十分了解的特性。尽管这种假设是推测性的，但如果你考虑一下这个事实：用力在眼睛上一按，这不属于

光学的刺激，却能引起视觉的一闪现（即“眼冒金星”），这表明这种推测有着某种可成立的根据。

贯穿全书的第二个问题是，对于把有意识的心灵视作像其他生物现象一样的一种生物现象，我们生来就具有一种文化上的抵御力。这可以回溯到17世纪的笛卡尔。笛卡尔把世界划分为两种实体：精神实体和物质实体。物质实体属于科学的专门领域，精神实体则是宗教的地盘。甚至到了今天，这种划分在某种程度上也被接受了下来。比如说，意识和主观性常常被看作是不适合于科学研究的题目。这种有意回避意识与主观性的态度，正是一种执着的客观化的倾向。人们往往认为，科学所研究的必须是关于客观地可观察的现象。有时，当我给生物学家和神经生理学家们讲课时，我发现，他们中的许多人很不愿意笼统地将心灵，特别是意识，划在科学研究的专门领域之内。

这第三个问题，不易让人意识到却贯穿在通篇内容中：我们用来讨论这些难题的传统术语在许多方面是不恰当的。就拿组成本书书名的心、脑与科学三个术语来说，只有“脑”一词是完全明确的。我所说的“心”仅仅是指：不论是有意识的还是无意识的、构成我们精神生活的思想、感情和经验的序列。但是，使用“心”一词，具有包含旧哲学理论中灵魂含义的危险。要想改变把心看成是一种东西、或至少是一个场所，或至少是所有心理过程在其中发生的某种黑箱的观点，是十分困难的。

“科学”一词的处境就更糟了。如果可能的话，我情愿不去使用这个词。“科学”在某种程度上已成为表示敬意的字眼，所有那些根本不同于物理学和化学的学科都渴望称自己为“科学”。我们可以记住这样一条好的经验：任何自称是“科学”的东西，很可能不是科学。例如，基督教科学、军事科学、也许还有认知科学或社会科学。“科学”一词，容易使人联想到许多身穿白

大褂的研究人员摇动着试管，注视着仪器。在许多人的心目中，“科学”意味着一种神秘的一贯正确性。我想提出如下与此相对的观点：我们在理智的学科中所极力追求的是知识与理解，不论在数学、文学批评、历史学、物理学或哲学领域里，我们所得到的也只是知识和理解。由于一些学科比另外一些学科更系统些，我们可以为那些相对系统些的学科保留“科学”一词。

我感谢许多学生、同事和朋友对我准备里思讲座，包括广播稿和此书的出版，所给予的帮助。我特别要感谢的有：艾伦·科德、雷雅纳·卡里翁、斯蒂芬·戴维斯、休伯特·德赖弗斯、沃尔特·弗里曼、巴巴拉·霍兰、保罗·库伯、卡尔·普里布拉姆、冈瑟·斯滕特和瓦内萨·万。

英国广播公司给予了我特殊的关照。感谢演讲部主任乔治·费希尔的大力支持和我的节目制作人杰夫·迪汉的通力合作。我的妻子，达格玛·塞尔始终在协助我的工作，我由衷地感谢她，并将此书献给她。

一 心身问题

数千年来，人们在不断地理解他们与周围宇宙的关系。出于种种原因，今天的许多哲学家们却有意地回避这类大问题。然而，这些问题依然存在着。在本书中，我将向其中的一部分问题发出挑战。

现在，最大的难题就是：我们具有一幅关于我们人类自身的常识性画面，这幅画面与我们关于物理世界的整个“科学的”概念很难取得一致。我们将我们自己视作世界上有意识的、自由的、自觉的、理性的行为者，而科学告诉我们，在这个世界上只存在无知觉、无意义的物理粒子。那么，我们怎样才能够使这两种概念取得一致呢？例如，在仅仅包含着无意识的物理粒子的世界中，怎么会同时也包含着意识？一个机械性的宇宙怎么会包含有意向性的人类——能够向自己描绘世界的人类？一句话，一种本质上是无意义的世界怎么会包含着意义？

这类难题又涌入到其他更具有当代色彩的问题中：我们应如何解释计算机科学和人工智能方面的最新成果——在力图制造智能机方面所取得的成果？特别是，数字计算机能否向我们作出关于人心的正确描述？为什么社会科学通常没有为我们提供关于我们自身的洞见，而相比较而言，自然科学却为我们提供了关于周围自然界的真知灼见？我们所接受的对人们行为方式的普通常识性的说明与说明的科学方式是什么关系？

在这第一章里，我想深入到那个被许多哲学家视作全部难

题之首的难题中去：我们的心与周围宇宙的关系如何？我想你们一定会认出，这就是传统的心身问题或心脑问题。在当代，这个问题通常采取这样的形式：心与脑是如何相关的？

我相信，对于心身问题会有一个相当简单的解答，这种解答既与我们所具有的神经生理学知识相一致，又与我们关于心理状态——痛觉、信念、欲望等等的性质的常识的概念并行不悖。但是，在解答这一问题之前，我想要问：为什么心身问题看起来这么难以解决？为什么到头来在几个世纪的哲学和心理学领域中还存在着"心身问题"，却不存在像"消化与肠胃问题"这样一类问题？为什么心比其他生物现象看起来更神秘？

我认为，这个问题之所以难以解答，一部分原因是由于我们坚持用过了时的17世纪的词汇，来讨论这个20世纪的难题。我想起，当我还是一名本科生时，就不满于那种在心的哲学中显然通用的选择：你要么是一个一元论者，要么就是一个二元论者。如果你是一个一元论者，那么，你要么是一个唯物论者，要么就是一个唯心论者。如果你是一个唯物论者，那么，你要么是一个行为主义者，要么就是一个物理主义者，等等。以下我的目标之一，就是努力去摆脱这些陈旧范畴的束缚。我们注意到，关于"消化与肠胃问题"，没有人会感到必须要在一元论与二元论之间做出抉择。那么，为什么应该把这个问题与"心身问题"区别开来呢？

除了词汇的问题以外，还存在着一个或者说是一系列难题。自笛卡尔以来，心身问题采取了以下形式：我们怎样才能够说明那种显然是完全不同的两种东西之间的关系？一方面是像我们的思想、情感一样的心理事物；我们把它们看成是主观的、有意识的、非物质的。另一方面是物理事物；我们把它们看作是具有质量、空间广延以及与其他物理事物因果地相互作用的。大部

分试图解决心身问题的方法局限在否定以上其中一类东西的存在上，或者在某种程度上去贬低以上一类东西的地位。在我们的理智发展的过程中，面对自然科学取得的一定成就，出现贬低人心的地位的倾向是不足为怪的。因此，近年来时髦的多数唯物主义心理观，诸如行为主义、功能主义和物理主义对心的看法，用明确或不明确的方式，以否定我们通常所认为的那种心理的东西为结论。也就是说，他们否定了的的确确是我们内在所具有的主观的、有意识的心理状态，否定了这种状态像宇宙中任何其他东西一样真实、一样不可还原的观点。

那么，为什么他们会持有这种见解？为什么这么多的理论家最终都否定了心理现象具有内在的心理特性呢？我相信，如果我们回答了这个问题，就会明白为什么心身问题长期以来显得如此难以解答。

心理现象有四个特征，由于这四个特征，使得它们看起来不可能纳入我们关于世界是由物质构成的“科学”概念。正是由于这些特征，使得心身问题确实难解。这些如此令人困惑的特征，竟使许多哲学、心理学和人工智能领域中的思想家们说出一些关于人心的奇怪而又难以置信的话。

这些特征中最重要的是意识性。在我写作此书和你阅读它的时刻，我和你都是有意识的。世界中包含着这些有意识的心理状态和事件，这的确是一个普普通通的事实。但是，我们又很难弄清楚在纯粹的物理系统中怎么会存在着意识。意识这玩意儿是怎样发生的？比如说，我颅骨中灰色和白色的粘性物怎么会是有意识的。

我觉得，意识的存在好像应该使我们感到惊奇，因为设想不存在意识的宇宙是很容易的。但是如果你一旦这样去设想，那么你就会看到你已设想了一个毫无意义的宇宙。意识是人类特

定存在的中心事实，因为如果意识不存在，那么我们生存所凭借的全部其他人类的特定方面——语言、爱、幽默等等，都将不可能存在。此外，我还觉得当代的哲学和心理学这么轻视对意识的探讨真让人感到几分羞愧。

人心的第二个不易把握的特征就是哲学家和心理学家所谓"意向性"；由于具有这一特征，我们的心理状态就指向、涉及、关联到或针对外在世界的种种客体和事态，而不是针对心理状态自身。另外，需要说明的是，"意向性"不仅仅讲的是意向，还包括信念、欲望、期待、恐惧、爱、恨、渴望、厌恶、羞愧、骄傲、忿怒、欢乐和所有那些涉及到或关联到外在世界的（不论是有意识的还是无意识的）心理状态。看来，关于意向性的问题很像关于意识的问题。我头脑中的这一团东西怎样才能关联到外物？它怎样才能涉及到外物？正像所有其他的物质实在是由虚空中的原子组成一样，在头颅中的这块东西毕竟也是由"虚空中的原子"组成的。那么，大致地说，虚空中的原子怎样才能表现万物？

人心的第三个特征是心理状态的主观性，这种特征看来也难以与关于实在的科学概念相一致。别人无法感觉到我自己的疼痛，这一事实表明了这种主观性的存在。我以我的观点来看世界，而你从你的角度去观察。我意识到自我和我的内部心理状态，但这与他人的自我和心理状态截然不同。从17世纪以来，我们已经认识到实在必须是为所有健全的观察者所公认的东西，即实在必须是客观的。那么，我们怎样去协调主观心理现象的实在与完全客观的实在的科学概念呢？

最后，第四个难题是心理的因果性问题。作为一种常识，我们都能设想我们的思想、感情与我们的行为方式确有不同，它们确实对物理世界有某种因果的影响。比如说，我决定举起我的

胳膊，瞧，它就举起来了。但是，如果我们的思想和感情确属心理的东西，那么，这些心理的东西怎样才能去影响物理的东西？心理的东西怎么会产生一种物理的差异呢？我们是否应该认为我们的思想、感情会以某种方式在我们的大脑和其余神经系统中产生化学的结果？这样的事情是怎样发生的？我们是否应该认为思想环绕在神经轴突四周，或挠动着神经树突，或潜行在细胞壁内并侵扰着细胞核？

如果我们不承认在心脑之间存在着这种联系，那么我们不是等于承认了心是可有可无的观点了吗？这岂不是把人心看作像对波浪的运动不起什么作用的泡沫一样不重要吗？我想假设泡沫是有意识的，它们可能自己寻思："一天到晚，把波浪往沙滩上推上拽下的，这是一件多么费力的工作啊！"但我们知道，泡沫对波浪来说是根本不重要的。那么，为什么我们设想我们的心理生活要比物理实在的波浪之上的一堆泡沫更重要些呢？

人心的这四个特征：意识性、意向性、主观性和心理的因果性正是使心身问题显得难以解答的原因。我想指出，它们都是我们内心生活真实的特征。并不是每一种心理状态都同时具有这些特征，但是任何完满的对心和心身关系的解释必须考虑到这四种特征。如果你的理论得出否定其中任何一种特征的结论，那么，你一定已在某处犯下了一个错误。

为"解决心身问题"我要提出的第一个论点是：

全部心理现象，不论是有意识的或无意识的、视觉的或听觉的，不论是痛觉、触痒觉、或思想，乃至我们全部心理生活，都是由在脑中进行的过程产生的。

为了对这一过程有所了解，我们试图至少对一种心理状态的因果过程作某种详细描述。例如，我们来考察一下痛觉。当

然，随着我们关于脑如何工作的知识的增加，我们现在所提出的看法也许在一代人之后会显得古老陈旧了些。但是，我们所提出的观点，尽管在细节上发生了变化，在解释问题的方式上仍会保持其有效性。按当前流行的观点，痛觉信号由感觉神经末梢传送到脊髓至少要通过两类神经纤维：一类是专门用于传递刺痛觉的△A纤维；一类是专门用于传递灼痛和疼痛觉的C纤维。在脊髓中，这些感觉信号，经过一个被称为利骚厄道的区域，最后到达脊髓神经元。当这些感觉信号上升到脊柱时，它们通过两条彼此分开的通路进入脑：刺痛觉通路和灼痛觉通路。这两条通路都要经过丘脑，但刺痛觉随后局限于大脑的躯体感觉皮质中，而灼痛觉通路不仅把灼痛信号输入到躯体感觉皮质中去，而且最后将它们输入到下丘脑和脑底部的其他区域中去。由于存在这种区别，对我们来说确定一种刺痛觉比确定一种灼痛觉要容易得多——例如，我们可以相当精确地指出某人正在用大头针刺我皮肤的某个部位——而灼痛和疼痛觉可能会更令人痛苦，因为这些痛觉信号激活了神经系统中更多的区域。这种实际的痛觉既来自于对脑底部区域特别是对丘脑的刺激，又来自于对大脑躯体感觉皮质的刺激。

为了达到我们讨论的目的，我需要完全弄清楚的一个观点是：我们的痛觉由一系列活动引起，它们发端于分离的神经末梢，终止于丘脑和脑的其他部位。的确，就实际的感觉而言，中枢神经系统中的活动足以引起痛觉，这一点，我们既可以通过被截肢者所感到的幻肢痛觉而得知，也可以通过人为地刺激脑有关部位引起的痛觉而得知。我想指出的是，一般说来，痛觉是如此，心理现象也是如此。我们可以大致地说，有关我们心理生活的一切活动，我们全部的思想和感情，包括我们现在讨论的作为脑的组成部分的全部中枢神经系统的活动，都是由脑的内部

过程产生的。就心理状态的产生而言，决定性的过程发生在头脑内部，而不在于外部或周围神经的刺激。对这一点的论证是简单的。如果活动仅在中枢神经系统之外发生，而丝毫不涉及到脑的话，就不会存在心理活动。但是，如果正常的情况发生在大脑中，即使不存在外部刺激，也会产生心理活动。（顺便提一句，外科麻醉的工作原理告诉我们：麻醉可以阻止外部刺激对中枢神经系统发生相应的影响。）

但是如果说痛觉和其他心理现象是由脑过程产生的，那么人们可能会问：什么是痛觉？痛觉究竟是什么？好吧，就痛觉本身而言，显而易见的回答就是，痛觉是种种不令人愉快的感觉。但这样的回答是不能让我们满意的，因为它没有告诉我们怎样才能用我们关于世界的整个概念去说明痛觉现象。

但我还是认为这个问题的答案是显而易见的，只是需要作某些说明。针对我们的第一个论点，即痛觉和其他心理现象是由脑过程产生的，我们需要增加第二个论点：

痛觉和其他心理现象正是脑（也许还包括中枢神经系统的其他部分）的特征。

本章的基本目的之一就是去揭示这两个命题如何同时为真。怎么会同时出现脑产生心，心又正是脑的特征这两种情况？我以为，正是由于没有看到这两个命题怎么能够同时为真，致使心身问题长期以来得不到解决。这两种观念容易引起对不同层次问题的混淆。如果说心理和物理现象具有因果关系，那么，其中一种现象怎么会是另一种现象的特征呢？这不正暗示着心可以自生——那令人生畏的自因理论吗？这里，引起我们困惑的是一种对因果性的误解。我们很自然地会设想，每当A引起B时，必定会有两种分离的事件。一种被确认为原因，另一种被确认为结果；全部因果关系的功能就好像台球相互撞击的方式一

样。这种粗糙的心脑因果关系模式，迫使我们去接受某种二元论的观点；迫使我们去认为，物质领域中的"物理的"事件是非物质领域中的"心理的"事件产生的原因。在我看来，这是一个错误，而改正这个错误的方法就是提出一种更精致的因果观。为此，我先暂且回避一下心脑之间的关系问题，而去考察自然界中其他种类的因果关系。

在物理学中，小与大的级别常用系统中的微观与宏观特性来划分。举例来说，考虑一下我身旁的桌子或我面前的这杯水。每个物体都是由微观粒子构成的。这些微观粒子都具有分子层次、原子层次以及更深的亚原子层次的特征。同时，每个物体也具有像桌子的固体性、水的液体性、玻璃的透明性这样一些性质，这些性质属于物理系统表面的或总体的特征。这些表面或总体的特征，其原因，可以用微观层次上元素的活动来解释。例如，我面前桌子的固体性，可用构成桌子的分子所处的一定网状结构来解释。与此相似，水的液体性，用H_2O分子之间相互作用的性质来解释。微观层次上的元素活动能从因果上解释那些宏观的特征。

我要指出的是，这为我们解答令人困惑的心脑关系问题提供了一种理想而又普通的模式。对于液体性、固体性和透明性，我们可以毫无困难地设想这些表面特征是由微观层次上元素的活动导致的，同时，我们还承认这些表面现象正是该系统的特征。我认为，阐明这种观点最清楚的方法，就是指出，物体的表面特征既是由微观元素的活动导致的，同时又体现在由微观元素构成的系统中。这里虽然存在着一种因果关系，但是与此同时，这些表面特征正是其活动在微观层次上导致那些特性的那个系统的高层次特征。

为了反对这种观点，有人也许会说，液体性、固体性等与微

观结构的特征是同一的。比如说，我们可以通过分子排列的网状结构去定义固体性，正像热常常与分子运动的平均动能是同一的一样。在我看来，这种看法是正确的，但是，它决不能推翻我所作的分析。对一种物体的表述方式，由最初通过感性所接受的表面特征来定义到以后采用引起表面特征的微观结构来定义，是科学进步的一种标志。举个固体的例子来说，我面前的桌子是固体的，也就是按通常意义说，桌子是有硬度的，它可承受压力，可在上面放书本，像其他桌子一样，它不易被多数其他物体穿透，等等。这些是关于固体的常识性看法。但如果从科学角度看，人们可以把固体性定义为引起这些总体可观察特征的微观结构明显特征。因为，人们或者可以说，固体性就是分子系统的网状结构，例如，这样被定义的固体性就会产生耐触、耐压性。人们或者可以说，固体性由像坚硬性、耐压、耐触性这样高层次的特征构成，这些特征是由微观层次上元素的活动导致的。

如果我们将以上所讲的原则用于对心的研究，从脑功能导致心理状态的角度来说明心脑关系，在我看来是没有什么困难的。正像水的液体性既是由微观层次元素的活动导致的，同时又是体现在微观元素系统中的一种特征那样，正是在这种"由……导致"和"体现在……中"的意义上，心理现象既是由在神经元或模块水平上脑中进行的过程导致的，同时又是在由神经元组成的这个系统中体现出来的。正像我们对任何物质系统需要作出微观与宏观的区分那样，我们也需要对脑作出微观与宏观的区分。尽管我们可以说一个粒子系统其温度是10℃，或是一种固体，或是一种液体，但我们不能说某个粒子是固体的，或是液体的，或其温度是10℃。比如，我不能从一杯水中分离出一个分子然后说："这个分子是湿的。"

正是本着这种原则，关于一个特定的脑，我们所知道的也只不过可以说，"这个脑是有意识的，"或说："该脑正体验着干渴或疼痛"，而不能说，该脑中的某个神经元"正处在疼痛之中，正体验着干渴"。总之，尽管存在着关于脑在细节上如何工作这个巨大的经验上的谜，但这并不妨碍我们在逻辑上、哲学上或形而上学上根据我们十分熟悉的从周围自然界中得来的方法去说明心脑之间的关系。在世界上，一种现象的表面特征，既由一种微观结构所导致，又体现在这种结构中，这是最普通不过的事实了，而心脑所展现出的那种关系也不过是这种关系罢了。

现在，让我们回到我所提出的四个难题上来，也就是我们要想解决心身问题所面临的那四个难题。

首先，意识何以是可能的？

说明某物如何是可能的最好方式，就是去揭示它如何实际地存在。关于痛觉在实际上怎样由丘脑和感觉皮质中的神经生理过程产生，我们已作了一个大致的描述。那么，为什么仍有许多人对这种解答感到不满意呢？我看，通过将这个难题与科学史上一个更早的难题相类比，这种疑惑感就不难消除。在很长一段时间内，许多生物学家和哲学家以为，原则上在纯生物学领域内描述生命的存在是不可能的事情。他们认为，除了生物的过程以外，还必定存在某种其他的因素。为了给死的和惰性的物质带来生命，就必须假设有某种生命冲动存在。今天，我们很难体验到甚至在上一代人中还持续着的活力论与机械论之争的激烈程度，今天，那些争论的问题已不再为人所关注。为什么呢？我想，这倒不是因为机械论的胜利和活力论的失败，而是因为我们对生命有机过程的生物学特性有了更多的了解。一旦我们明白了应该怎样在生物学上对这些生物特性进行解释，活物

之谜对我们来说就不再成其为谜了。我看，我们应该将类似的思路引伸到我们对意识的讨论上来。从原则上讲，这块物质，即这块有意识的灰白色麦片粥状的脑质，不应该比从前那块似乎是神秘的物质，即那块有生命的、粘挂在一种钙架上的核蛋白分子聚集体更显得神秘。一句话，消除这种神秘感的途径，就是去理解脑过程。对这些过程我们至今仍不十分清楚，但我们了解脑过程的一般特性，我们知道在神经元或神经元模块中存在着某些特定的电化学活动，也许，脑的其他特征与这些过程一道引起了意识现象。

我们的第二个难题是，真空中的原子怎么会具有意向性？它们怎么能涉及到外物？

正像我们对待第一个问题一样，说明某物是如何可能的最好方式，就是去揭示其怎样实际地存在着。让我们来考察一下渴觉。据我所知，至少某些种类的渴觉是在下丘脑中由连续的神经传导引起的。这些连续的传导依次由下丘脑中血管紧张肽的活动引起，而血管紧张肽依次由肾脏分泌出的高血压蛋白原酶合成。渴觉，至少这些种类的渴觉，是由发生在中枢神经系统中，主要集中于下丘脑中的一系列活动引起的，并在下丘脑中体现出来。感到口渴起码就是有喝的欲望。所以，渴觉是一种有意向的状态：它具有内容；它的内容取决于满足它的条件。渴觉具有所有其他意向状态所共有的特征。

正像对待生命与意识之"谜"一样，解开意向性之谜的方法就是要尽可能详细地去描述那些怎样由生物的过程引起，同时又体现在生物系统中的现象。视觉和听觉、触摸觉、饥饿感、渴觉以及性欲都是由脑过程引起的，并体现在脑结构之中，它们都是具有意向的现象。

我可不是说，我们将会摆脱对自然的神秘感。相反，我的

这些例子都是在一种使人感到惊讶的意义上引用的。但是，我要指出的是，它们比起世界上其他令人惊讶的特征，诸如万有引力的存在、光合作用的过程以及银河的边际，其神秘性不多也不少。

我们的第三个难题是，我们怎样将心理状态的主观性纳入到一种关于现实世界的客观性概念中去？

在我看来，认为实在的定义应排除主观性，这是个错误。如果“科学”一词是指我们关于世界的客观和系统的真理之集合，那么，主观性的存在就像其他事实一样，也是一种客观科学的事实。如果一种关于世界的科学陈述试图揭示事物的真相，那么这种陈述的特征之一就是心理状态的主观性。因为生物的进化产生了某些种类的生物系统，即具有主观特性的人与某些动物的脑，这恰恰是关于生物进化的一个普通事实。我此时的意识状态是我脑的一个特征，但这种对我是可以理解的有意识的方面对于你倒是不可理解的。同样，你当下的意识状态是你脑的一个特征，它的有意识的方面对于你是可以理解的，对我倒是不可理解的了。所以说主观的存在是一个生物学的客观事实。利用现行科学理论的某些特点去定义“科学”是一种一贯的错误。一旦这种狭隘的观点被视作偏见，那么，任何事实的领域都应被视作系统化研究的主题。比如说，如果上帝存在，那么这也是一个同其他事实一样的事实。我不知道上帝是否存在，但我毫不怀疑主观心理状态的存在，因为我现在正处在这种状态，而你也如此。如果主观性的事实与某个“科学”的定义相悖，那么，我们不得不放弃的是这个定义而不是那个事实。

第四，关于心理的因果性问题，我们现在的目的是要解释心理活动怎么会引起物理活动的。例如，像思想那样的“毫无重量”、“虚无缥缈”的一种东西，怎么会导致一种行动？

回答是，思想并不是没有重量和虚无缥缈的东西。当你具有一种思想时，脑活动实际上在进行着。脑活动通过生理过程引起身体的运动。因为心理状态是脑的特征，所以对心理状态就有两种层次上的描述：一种是较高层次的使用心理术语的描述；一种是较低层次的使用生理术语的描述。脑系统中完全一样的因果力可以分别在这两种不同的层次上进行描述。

我们可以再一次使用一个物理学的类比来说明这些关系。考虑一下锤子钉钉子的情况，锤子和钉子都具有某种固体的性质。用棉花和黄油制作的锤子根本起不到锤子的作用，用水或蒸汽制作的锤子根本就不叫锤子。所以，固体性是锤子的一个真实的因果特性。但是，固体性自身又是由微观层次上粒子的活动导致的，并体现在由微观元素组成的系统中。脑中有两种真实的因果关系描述层次，一种是心理过程的宏观层次；另一种是神经元过程的微观层次，这两种过程恰恰类似于锤子的那两种真实的因果关系描述层次。例如，意识是脑的一个真实的特性，它能导致事情的发生。我要完成一个像举起手臂这样的动作，我的这种有意识的企图，导致了我的手臂的运动。从较高的那种描述层次来看，这是举起我的手臂的意向引起了手臂的运动。但从较低的那种描述层次来看，这是由一系列神经元放电导致肌肉收缩的一连串活动引起的。正像用锤子钉钉子的情况一样，对同一序列的活动，可在两种层次上加以描述。从因果关系上看，这两种层次的描述都是真实的，那些较高层次的因果特征，既由较低层次元素的结构引起，又体现在那种结构之中。

总之，在我看来，心与身是相互作用的，但它们不是两种不同的东西，因为心理现象就是脑的特征。了解这种观点的一个途径，就是既把它看作是物理主义的主张，又把它看作是心理主

义的主张。我们把"朴素物理主义"的观点表述为，世界上所有存在物都是具有性质和关系的物理微粒。这种实在的物理模式其影响之大，以致我们难以发现我们怎样才能对朴素物理主义提出有分量的质疑。我们把"朴素心理主义"的观点明确表述为：心理现象是真实的存在。确实存在着心理状态；其中一部分是有意识的；大部分是具有意向性的；全部心理状态都是具有主观性的；大部分心理状态在决定世界中物理事件时起着因果作用。至此，第一章中的论点可以十分简明地表述出来了。朴素心理主义与朴素物理主义彼此是完全一致的。实际上，就我们对世界如何运作的所知而言，它们不仅是一致的，而且两者皆为真。

二　计算机能思维吗？

在前一章，我至少已勾出了解决所谓“心身问题”的轮廓。尽管我们对脑功能的细节还不了解，但我们已明白了脑过程与心理过程的一般关系。心理过程是由脑要素的行为活动产生的，同时又体现在由这些要素组成的结构中。我认为，这种对“心身问题”的回答与关于生物现象的一般生物学观点是相一致的。如果我们以我们关于世界如何运作的已有知识为前提，那么这种回答的确是一种常识性的回答。可是，这种观点仅仅是少数人的见解。目前盛行于哲学、心理学以及人工智能领域中的观点是强调人脑功能与数字计算机功能相似的观点。在这类观点中有一种最极端的看法，认为人脑不过是一台数字计算机，人心不过是一种计算机程序。人们可以把这种观点概括为：心与脑的关系就是程序与计算机硬件的关系。我把这种观点称为“强人工智能”观点，简称“强AI”观点。

按照这种观点，人心根本谈不到什么生物学的性质。脑不过是无限多种不同硬件计算机的一种，这种硬件计算机能够装载构成人类智能的那些程序。按照这种观点，任何具有合适的输入与输出程序的物理系统，都与你我一样，在完全同等意义上，具有心灵。这就好比说，一台用旧啤酒罐做成的、以风车为动力的计算机，只要它具备适当的程序，就应该有其心。这种观点并不是认为计算机说不定会具有思想与感情，而索性认为计算机肯定会具有思想与感情，因为从这种观点来看，只要有了合

适的程序，计算机就会有思想和感情。

多数持有这种见解的人们认为，我们目前还不能设计出心理的程序。但他们当中相当普遍的意见认为，计算机科学家和人工智能领域中的工作者们设计出与人脑和心灵相当的合适的硬件与程序仅仅是一个时间的问题。这些硬件与程序作为人工的脑与心将在各个方面都和人脑与人心相同。

许多非属人工智能领域的人们很难理解为什么竟有人会相信这样的观点。为此，在对这种观点进行批驳之前，我先举几个例子，让大家看看搞这一行的人们实际是怎样说的。卡内基-梅隆大学的赫伯特·西蒙说，我们已经拥有真正能够思想的机器，根本不需要去等待某种未来机器的出现，因为眼下的数字计算机已经具有与你我完全同等意义上的思想了。请想一想吧！哲学家们数世纪以来为弄清楚机器能否思维伤透了脑筋，今天，我们发现在卡内基-梅隆大学里已经有了这种机器。西蒙的同事艾伦·纽厄尔声称，我们现已发现（请注意纽厄尔说的是“发现”而不是“假设”或“考虑到有可能性”，他说的是已经发现），智能不过是一种物理符号处理的问题；它与任何特定类型的生物或物理的湿件或硬件没有本质的联系。而且，凡是能够以恰当方式处理物理符号的系统，都具有与人类智能完全同等意义上的智能。值得称赞的是，西蒙与纽厄尔都强调他们的这些主张决不含有任何隐喻的成分，他们的话是说一不二的。弗里曼·戴森说过，计算机一旦能够进化，就会胜过它周围的人类。因为意识不过是一种“形式化过程”，在计算机中，这些形式化过程存在于计算机的物质形态之中，这种物质形态在逐渐冷却下来的宇宙中能更好地存活下来，而远胜潮湿而浑浊的人体。马萨诸塞理工学院的马文·明斯基说，下一代计算机将会十分聪明，到那时，“如果它们愿意把我们收进它们的家门，作为家饲的爱畜，我

们就算是幸运的了。"在夸大数字计算机能力的文献中，我最欣赏"人工智能"一词的创造者约翰·麦卡锡的话。他说，甚至"像恒温器那样简单的机器都可以说具有信念"。的确，在他看来，几乎所有能够解决问题的机器都可以说具有信念。我佩服麦卡锡的勇气，我曾经问过他："你的恒温器具有的是什么信念？"他回答说："我的恒温器具有三种信念——这里太热，这里太冷，这里温度适宜。"作为一个哲学家，我欣赏所有这些主张。理由很简单，与多数哲学论点不同，这些主张很明确，对它们的反驳也只要简单明确。在这一章中，我就来反驳这些论点。

这个反驳，与任何特殊的计算机技术阶段无关。强调这一点很重要，因为，这里容易产生一种误解，即认为要想解决我们的难题必须期待着某种尚未创造的技术奇迹。而事实上，我们的这个反驳是完全独立于任何技术阶段的。它只同一台数字计算机的那个定义有关，只同什么是一台数字计算机有关。

数字计算机概念的要点是其运算能被纯形式化地表述。这就是说，我们是用抽象的符号，比如打印在带子上的0与1的符号串，来表述计算机运算过程中的步骤的。一种典型的计算机"规则"决定一台机器处理某种状态、其带子上有某种符号时，执行某一运算，例如擦掉一个符号，或打上另一个符号，然后进入另一状态，例如让带子往左移一格。但是符号并没有意义，它们不具有语义内容；它们不涉及任何事物。它们只能纯粹按照形式的或语法的结构加以确定。比如上面讲的0与1只不过是数码，甚至不能代表数。事实上，数字计算机正是由于这一特点，才会具有如此巨大的能力。同一类硬件，如果设计得当，能够运算无数种不同的程序。同一个程序，能够在无数种不同类型的硬件上使用。

但是，程序具有的这种被纯形式化、语法化地定义的特征，对

于那种把心理过程与程序过程视为同样过程的观点是致命的。理由十分简单，人心不仅仅是形式的或语法的过程。根据定义，我们的内在心理状态具有某种内容。比如，我现在想起堪萨斯城，或者想喝一杯啤酒，或者担心利率是否会下跌，在这每一种状态下，我的心理状态除了具有各种可能的形式特征以外，还具有某种心理内容。即使思想以一连串的符号的形式出现在头脑中，但是思想所包含的决不止一些抽象的符号串，因为符号串本身没有任何意义。如果我的思想要涉及任何事物，符号串也必有某种意义，使得思想能够涉及到某些事物。一句话，人心不仅是语法的，它还有一个语义的方面。计算机程序永不可能替代人心，其理由很简单：计算机程序只是语法的，而心不仅仅是语法的。心是语义的，就是说，人心不仅仅是一个形式结构，它是有内容的。

为了说明这个观点，我设计了一种想象的试验。设想一些计算机程序设计家们设计出一种能够使一台计算机模拟“理解”汉语的程序。例如，如果用汉语向这台计算机提一个问题，它会就这一问题运用自己的存储器或数据库，对这个问题给出合适的汉语的回答。为了便于讨论，假设这台计算机的答案与一个地道的中国人的答案一样。那么，根据这种情况，能否说这台计算机已理解了汉语？能否说它完全和中国人一样地理解汉语？假设你被关在一间屋子里，屋内有满满几筐汉语符号，假如你（与我一样）对汉语一字不识，但给你一本用英语写的用来处理这些汉语符号的规则，这些规则是按照汉语的语法而不是语义，对符号的处理加以纯形式的规定的。比如，规则中可能有这样的话：“从一号筐中取出‘甲’符号，将它放在从二号筐中取出的‘乙’符号之后。”假设这时又有另外一些汉语符号被送进屋来，你根据相应的规则将汉语符号送出屋外。假如你根本就不

知道送进屋来的这些符号就是屋外人的"问题"，你也不知道你送出屋外的那些符号就是所谓"问题的答案"。此外，假设那些程序设计家们善于设计程序，而且你又善于处理符号，这样，你的答案就同一个地道的中国人作出的回答没有什么不同。你被这样关在屋子里，将你的汉字符号移来移去，用送出去的汉字符号回答送进来的汉字符号。按照我描绘的这种情况，你仅仅是摆弄这些形式符号，不可能学到哪怕一丁点儿汉语。

以上的观点不过就是说，在一个外部观察者看来，依靠使用一种形式的计算机程序，虽然你操作得准确无误，好像你已理解了汉语，但你仍旧是目不识丁。如果说通过适当的计算机程序来获得对汉语的理解力对你来说是不够的话，那么，对任何数字计算机来说也是不够的。同样，这个理由说起来也很简单。如果你不懂得汉语，那么计算机也不会懂汉语，因为数字计算机不可能仅仅依靠一种程序就会具有你所没有的东西。计算机所具有的，也正像你所具有的，是一种处理未经翻译的汉字符号的形式程序。也就是说，计算机具有的是一种语法，而不是语义。"汉语屋子"这个比喻的全部用意就在于提醒我们注意到一个为我们所熟知的事实。要理解一种语言，以至要完全具有心理状态，就要具备比仅仅一套形式符号更多的东西，就要具备一种释义，或者说那些符号都要有意义。而每一台数字计算机正像我们所定义的那样，不会比单纯的形式符号具有更多的东西。因为正像上文所说，计算机的运算是按其执行程序的能力来定义的。而这些程序通过纯粹形式就可加以规定，也就是说，它们不具有语义内容。

如果我们对比一下，用英语问答是什么样的，用某种我们一窍不通的语言问答又是什么样的，我们就会明白上述论证的力量。设想你在那个汉语屋子里也接到一些用英语提出的有关

你的年龄或经历的问题，并要求你回答。那么，使用汉语与英语这两种情况有什么不同呢？我们再假设，如果你同我一样不懂汉语而懂英语，那么这种不同就是显而易见的。你之所以理解英语的问题，是因为这些问题是用你能理解其意义的符号表达出来的。同样，当你用英语作出回答，你也就给出了对你具有意义的符号。而当你用汉语处理问题时，则全然不是这么回事。在这种情况下，你不过是在按照一套计算机程序处理着形式的符号，你不会给任何因素添上任何意义。

对于这个论证，人工智能与心理学界的研究者们，或哲学家们有各式各样的回答。这些回答有一个共同点：都没有充分的理由。这些回答必然没有充分的理由，原因很明显：我们的这个论证是建立在一种非常简单的逻辑真理之上的，即，单纯的语法对语义来说是不充足的。数字计算机既然是计算机，按照定义，就只有一套语法而已。

为了使这一论点更加清楚，我想再谈谈两种经常出现的反对意见。

有人企图用"计算机系统的整体是理解汉语的"这种说法来反驳"汉语屋子"的比喻。他们认为，尽管我这个在屋内处理符号的人不懂汉语，但我只是计算机系统中的加工中心的部件。他们认为，可以把屋子、满筐满筐的符号和规定程序的册子，也许还有其他组成部分，看作一个整体，而这个整体是能够理解汉语的。这种观点正是我在上文中已经驳斥过的观点。这样的系统没法从语法中得出语义。我，作为符号加工的中心部件，既然无法领会任何一个符号的意义，那么整个系统也同样如此。

对我所作的比喻提出的另一种常见的答复，是设想将理解汉语的程序装进一个机器人中。如果这个机器人按程序行动起来，并与外界发生相互的因果作用，这不是足以说明机器人已理

要回答这个问题，又取决于你所说的这种人造物是什么。假设我们设计出一台与一个活人的每一个分子都一样的机器，那么你能复制其作为原因的身体，你也就大致能复制其作为结果的心灵了。因此，对这个问题，至少在原则上仍然可以不假思索地回答说是能思维的。如果你能制造一台与一个活人同样结构的机器，那么这台机器大概是能够思维的。甚至可以说，它将成为一名人类的接班人。好了，让我们再重新考虑这个问题。

我们的问题并非："一台机器能否思维"或者"一件人造物能否思维"，我们的问题是"一台数字计算机能否思维"。对这个问题我们仍要非常谨慎地加以解释。从一种数学的观点来看，任何东西都可以被描述成仿佛是一台数字计算机。那是因为任何东西都能被看作正在表示或执行一个计算机程序。从一种全然空泛的意义上说，我面前桌子上的一枝钢笔也可以被看作是一台数字计算机。它正好有一个十分单调的计算机程序。这个程序说："呆在那里。"既然任何东西都可被看作是在执行一个计算机程序，在这种意义上，任何东西都可以算作一台数字计算机，这样我们的问题就再一次得出了一个没有什么意义的答案。由于人脑能执行多种计算机程序，人脑当然就是数字计算机，而且人脑当然是会思想的。所以，这又是一个意义不大的答案，而且它回答的根本不是我们所问的那个问题。我们所要问的问题是："像我们所定义的那种数字计算机能否思维？"也就是："实现具有合适输入与输出的适当的计算机程序，是否可以算作思想，是否构成思维？"对这个问题的回答就与对前几个问题的回答不同，其答案是明确的："不能。"说它不能思维，其理由我们早已讲清楚，即，计算机程序是按照纯语法加以定义的。思维绝不仅仅是处理无意义的符号的问题，它包含有意义的语义内容。这些语义内容就是我们所说的"意义"。

再强调一下我们的讨论与计算机技术发展到什么阶段无关，这一点很重要。我们的论证与即将到来的、令人惊叹的计算机科学的进展无关。它与串行处理与并行处理的区分无关、与程序的规模大小无关、与计算机运算的速度无关、与那些能主动地与它们周围环境相互作用的计算机无关、甚至与机器人的发明无关。技术的进步总被过分地夸大，但即使去掉这些夸张的成份，计算机的发展也是很可观的，我们有理由期待它在未来取得更显著的进展。毫无疑问，我们将比今天，当然也比过去更善于用计算机来模拟人的行为。但是，我现在阐明的是，如果我们谈的是心理状态或有一颗心的问题，那么任何这类模拟都与我们谈的内容不相干。不管技术有多么高超，不管计算机的运算速度有多快，都与我们的问题无关。如果心当真是一台计算机，它的运算必须用语法来定义，然而，意识、思想、感情、情绪以及心理的所有其他特征远非语法所能包容。不管计算机的模拟能力有多强，按照定义，它也不能复制那些特征。这里的关键性区别就是复制与模拟的区别。就模拟自身而言，它永远不能构成复制。

至此，我所做的不过是提出一种根据，用以说明我在本文开始时所引的那些言论确实就像看上去那么荒谬。其实，在这个讨论过程中存在着一个令人费解的问题，这个问题是："为什么人们总认为计算机能思维或有感情、情绪以及心理的其他特征呢？"当然，我们能用计算机来模拟任何形式化的过程。比如，我们能用计算机来模拟英国的货币流通量或工党力量分布的形势。我们能用计算机来模拟伦敦附近各郡的暴雨或伦敦东区仓库的火灾。但是，在每一种情况下，没有人会把计算机的模拟看作是事实本身。没有人会想象计算机模拟的一场暴风雨会将我们浑身上下浇得湿淋淋。也没有人会想象计算机模拟一场火灾

会把房屋烧掉。那么作为理智正常的人究竟为什么会认为对一种心理过程的计算机模拟就是心理过程本身呢？我确实不知如何回答这个问题，坦率地说，在我看来这种想法一开始就是荒唐的。但对此我可以作两种推测。

首先，在对心的看法上，许多人仍被某种行为主义观点所作弄。他们认为如果一个系统就其行为来说好像是理解了中文，那么它就必然真正理解了汉语。但是，在讨论"汉语屋子"的比喻时，我们已经反驳了这种形式的行为主义见解。还有许多人主张的一种看法是，心灵不是生物世界中的一部分，不是自然界的一部分。"强人工智能论"的基本观点是，心是纯形式的；这种观点大致认为，心不能像其他生物产物那样被当作生物过程的一种具体产物。简言之，这是一种残余二元论。人工智能的信徒们相信心灵不仅仅是自然生物界的一部分；他们相信心是能被纯形式地表述的。不过，这就出现了一个悖论：在人工智能的文献中充斥着对某种所谓"二元论"的斥责，但事实上，整个"强人工智能论"却是建立在一种二元论的基础之上的。它拒斥这样的观念，心与其他生物现象一样，不过是一种自然界的生物学现象。

我想通过把这一章的论点与上章的论点结合在一起来结束本章的讨论。这两章的论点能非常简明地表述出来，的确，我的表述也许过于笼统，但如果我们把这些观点放到一起，我想我们会得到一种强有力的关于心、脑与计算机关系的概念。我的论证有一个十分简单的逻辑结构，便于读者判断论证是否有效。第一个前提是：

1. 脑产生心

当然，这未免太笼统了。它的意思是说，那个我们认为构成

心的心理过程，完全是脑内部进行的过程所产生的。我们把这点压缩成三个词——脑产生心。这正是世界真实状况的一个事实。第二个前提是：

2. 语法不足以满足语义

此命题是一个概念真理，它只是明确了我们对什么是纯形式的和什么是有内容的概念的区分。现在在这两个命题，即，脑产生心和语法不足以满足语义之外，我们再增加第三和第四个命题：

3. 计算机程序是完全以它们的形式的或语法的结构来定义的

我把此命题作为依定义而真的命题；它是我们所说的计算机程序概念中的一部分。

4. 心具有心理的内容，具体说具有语义内容

此命题我认为仅是关于我们心理活动的一个明显事实。我的思想、信念和欲望对它物有所涉及、有所关联，或者说它们涉及世界上的事态；其所以如此，是因为它们的内容使它们指向世界上的这些事态。现在，我们从这四个前提能得出我们的第一个结论，这个结论很明显是从第2、3、4个前提引出的：

结论 1 任何计算机程序自身不足以使一个系统具有一个心灵。简言之，程序不是心灵，它们自身不足以构成心灵。

这是一个强有力的结论，因为它意味着想仅仅通过程序设计来创造心的工程从一开始就注定要失败。创造心与任何特定的技术水平和程序的复杂性状态毫无关系，这一点再强调一次也是很重要的。这是一个纯粹形式的或纯逻辑的结论，它是从所有（或几乎所有）有关的论辩者公认的一组公理引出的。作为一个生物学的事实，脑过程产生心理状态，这一点甚至大多数"人工智能"观点的铁杆分子也同意。他们也承认程序是纯形式

地加以定义的。但是，如果将上述结论与我们已有的某些其他知识结合起来，那么立即就能看到，"强人工智能论"的目标是不可能实现的。

有了这些公理，我们再来看看由此还能导出些什么。以下是第二个结论：

结论2 脑功能产生心的方式不能是一种单纯操作计算机程序的方式。

这第二个结论来自第一个前提与第一个结论的结合。这就是说，我们从脑产生心、程序不足以构成心这两个事实中得出了这个结论，脑产生心的方式不能是一种单纯操作计算机程序的方式。我认为这也是一个重要的结论，因为这一结论表明，脑不是或至少不仅仅是一台数字计算机。在上文中，我们说过任何东西都可以被简单地看作一台数字计算机，脑也不例外。现在这个结论的重要性就在于以脑在计算方面的性能来解释脑产生心理状态的功能是根本不够的。的确，这个结论对我们来说，应该算是一个常识性的科学结论，因为它只是提醒我们：人脑是生物性的机器，关键在于人脑的生物学性质。这就与人工智能领域中有些人所主张的不一样，心出现在人脑中这一事实，不是一个与心的性质无关的事实。

从我们的第一个前提，我们还能得到第三个结论：

结论3 任何其他事物，如要产生心，应至少具有相当于脑产生心的那些能力。

这是从第一个前提中得出的一个简单结论。这有点类似说，如果我的汽油发动机能使我的车子每小时行驶75英里，那么任何能达到这种速度的内燃发动机就应至少具有相当于我的汽油发动机的输出功率。当然，某种别的什么系统，可能会通过使用与脑实际上所使用的完全不同的化学或生物化学特性来造成心

理过程。在其他星球或星系可能存在着生物，他们可能具有心理状态，并且可能依靠一种与我们完全不同的生物化学。设想火星人来到地球，而且我们断定了他们是有心理活动的。假如把他们的头颅打开，我们发现其中都是些绿色的粘液。如果这些绿色粘液的功能能产生意识以及他们的其他精神生活，那么这些绿色的粘液仍需具有相当于人脑产生心的能力。现在，从我们的第一个结论：程序是不够的，以及我们的第三个结论：任何其他产生心的系统必须具有相当于人脑产生心的能力，马上就可以推出第四个结论：

结论4.对于任何我们可能制作的、具有相当于人的心理状态的人造物说来，单凭一个计算机程序的运算是不够的。这种人造物必须具有相当于人脑的能力。

我相信，我们讨论的结果会使我们想起我们所熟知的事：心理状态是生物学的现象。意识、意向性、主观性和心理的因果性与生长、繁殖、胆汁分泌、消化一样，都是生物生命史的组成部分。

三 认知科学

我们可以十分自信地说这样的话："因为巴兹尔欣赏撒切尔夫人对福克兰群岛事件的处理，所以他投票赞成保守党。"但我们却没有把握说这样的话："由于巴兹尔的下丘脑处于一种状态，所以他投票赞成保守党。"也就是说，我们能运用心理方面的词汇，运用有关人们的欲望、愿望、恐惧和希望等方面的词汇，对人们的行为作常识性的说明。我们设想一定还存在着利用人们脑过程方面的词汇对人们行为所作的一种神经生理学的说明。而困难就在于，第一类说明在实际应用方面是行之有效的，然而不是科学的；相反，第二类说明当然是科学的，但是在实际当中我们不知道怎样使用它们。

因此，在我们面前显然有一条鸿沟——脑与心之间的鸿沟。20世纪一些最富有理智的探索一直在为填平这条鸿沟而努力，由此产生了一门关于人的行为的科学，它既不是常识性的传统心理学，也不是科学的神经生理学。时至今日，为填平这条鸿沟所作出的种种努力毫无例外地遭到了失败。其中行为主义的失败最为引人注目，但在我的有生之年，我还耳闻目睹了那些代表对策论、控制论、信息论、结构主义、社会生物学以及种种其他理论的不切实际的主张，而这些不切实际的主张最终是要落空的。先把话说在前头，我认为所有那些要填平这条鸿沟的努力都将是徒劳的，因为根本就没有什么鸿沟可填。

最新近的填平这条鸿沟的努力所依据的是人类与数字计算

机之间的相似性。按照这种观点中最极端的看法，即那种被我称为“强人工智能论”或“强AI”的看法：脑是一台数字计算机，心仅仅是一套计算机程序。对这种观点我在上一章已予以驳斥。与此相关，最近还有一种试图填平这条鸿沟的观点，由于它来源于认知心理学与人工智能的研究成果，同时又构成了“认知科学”这一新学科的主流，所以人们常把这种观点称为：“认知主义。”同“强人工智能论”一样，它认为计算机是心灵的真实图式，而不仅仅是对心灵的一种比喻。但它又与强人工智能的观点不同，它没有主张，至少不必要主张，计算机确实具有思想与感情。

如果要对认知主义的研究纲领加以概括的话，那么可以这样讲：思想是处理信息的过程，但是，信息处理不过是符号操作。计算机的工作是符号操作，所以，研究思想（他们宁愿把思想称为“认知”）的最好方法，就是去研究不论是在计算机中的或是在脑中的计算方面的符号操作程序。按照这种观点，认知科学的任务是去描述脑的特征，但它不是在神经细胞层次上，也不是在有意识的心理状态层次上，而是在把脑功能作为一种信息处理系统的层次上去描述脑特性的。这正是心脑之间的鸿沟可以得到填补的地方。

这项研究工程在心的科学中似乎是一次重大突破，但我不能夸大这种程度。的确，在认知主义的支持者们看来，既然认知主义已冲破了行为主义的迷雾，就可以说它是一次名符其实的突破，是一次将最终把心理学建立在安全可靠的科学基点上的突破。

在这一讲，我将向认知主义提出挑战，但我想先说明一下认知主义本身具有的迷惑力。我们知道存在着一个朴素的、常识性的、传统的心理学层次，也存在着一个神经生理学层次，即一

个由神经元、神经元模块、突触、神经介质、小结以及所有其他部分组成的层次。那么，为什么有人会认为在这两种层次之间还存在着一个属于计算处理的心理过程层次呢？为什么有人会认为，在种种层次上，脑所执行的那些我们视为机体生存本质的功能就是信息处理的功能呢？

其原因有几条：首先我要提到的是一个多少有点儿不体面的原因，但我认为它确实非常有影响。这就是：由于我们对脑还没有充分地理解，我们经常沉湎于使用最新的技术来构造一种试图理解脑的模式。在我的孩提时代，我们总是相信脑就是一部电话交换机。（“它还能够是什么呢？”）有趣的是，英国伟大的神经生理学家谢灵顿认为，脑的运转与一个电报系统是一样的。弗洛伊德常常把脑比作液压和电磁系统。而莱布尼茨则把脑比作一台碾磨机。据我所知，一些古希腊人以为脑是像一架石弩那样运转的。而今，很显然，关于脑的这个比喻就是数字计算机。

顺便指出，这种情况迎合了今天我们所听到的那些关于计算机和机器人的大话。流行的新闻报道常常使我们相信，我们不久就会拥有能干所有家务事的家庭机器人了，它们能为我们带孩子，能以其活泼的谈吐取悦于我们，还能在我们的晚年照顾我们。这当然是一派胡言。我们还远远不能制造能做其中任何一项事情的机器人。的确，那些成功的机器人，只是在非常有限的关系当中，比如像在自动生产流水线中那样，限于做非常固定的工作。

现在，让我们回到人们相信认知主义为真的那些郑重的原因上来吧。首先，他们认为他们确实有某些真实的心理学的证据。有两个证据。第一个证据来自反应-时间试验，这一实验表明，人们完成不同的智力劳动要花费不同的时间量。作为一个

证据，其思想是：如果人们所花时间量的差与一台计算机所需时间量的差是相对应的话，那么，这就至少证明了人的系统与一台计算机在同样的原理下工作。第二种证据来自语言学。尤其来自乔姆斯基和他的同事们关于生成语法的研究成果。作为一个证据，其思想是：讲一种语言所遵循的语法的形式规则与一台计算机所遵循的形式规则是一回事。

关于那个反应-时间的证据，我不想多说，因为我想大家都会认为，这个证据根本就不具有说服力，并且对它可以有许多不同的解释。我想要讨论的是关于语言学的那个证据。

可是，在对这两种证据的计算解释背后还有其更深的依据，我相信这也是对人们接受认知主义产生更大影响的一个原因。人们之所以接受认知主义的第二个原因就是以上两种证据所要例证的一个总的论点，它是这样一种观点：由于我们能够设计在处理信息时遵循规则的计算机，又显然由于人类在思考时，也遵循着规则，所以，这就存在着某种一致性，即人脑与计算机两者之间在功能上存在着一种相似的、甚至可能是相同的方式。

在认知主义研究纲领背后的第三个假设，是一个由来已久的观点。这种观点可回溯到莱布尼茨，甚至柏拉图。它假设一种心理的成果必有其理论的原因。它假设如果一个系统的输出是富有意义的话，例如，从这个意义上说，我们学习一种语言的能力或辨认他人面孔的能力是一种富有意义的认知能力，那么，就必定存在着某种理论，以某种方式内在化于我们的脑中，构成我们这种认知能力的基础。

最后，人们之所以追随认知主义的研究纲领，尤其是在哲学上倾向于认知主义，还有另一个原因，这就是他们找不到任何其他方式去理解心脑之间的关系。既然我们理解了计算机程序与计算机硬件的关系，它也就提供了一种极好的，也许是唯一的能

使我们说明心脑之间关系的模式。对于这种观点我在第一章已予以回答，本章毋须赘述。

我们将如何看待这些为认知主义提供的论据呢？对于认知主义，我不大相信自己也能像对待强人工智能的观点那样，提出一种置对方于死地的反驳。但是，我坚信如果我们去考察那些支持认知主义的论据，我们将会看到这些观点是极其软弱无力的。揭露了它们的弱点，我们就能够理解人类行为与计算机功能之间的几种重要的区别。

让我们先从“遵守规则”这个概念讲起。我们知道人类遵守着规则，计算机也遵守着规则。但我想指出，两者存在着一种根本性的区别。就人类方面来说，每当我们遵守一条规则，我们就在接受某种实际内容或规则意义的指导。就人类遵守规则而言，意义导致了行为。当然，仅仅通过意义还不能导致行为，但它们无疑对行为的产生起着一种因果作用。例如，考虑这样一条规则：在英国，车辆需沿道路左侧行驶。所以，每当我来到英国，我就要叮嘱自己记住这条规则。这说明什么呢？说明我服从了这条规则，也就等于说这条规则的意义，即它的语义内容对我的实际行为起着某种因果作用。请注意，还有许多其他总对所发生的事情进行描述的规则。但它们不是我正在遵守的规则。比如，假设我在一条双行道上行驶，方向盘位于我车子的右侧，那么你可以说我的行为符合下述规则：让方向盘沿道路中心线行驶。这的确是对我行为的一个正确描述。但它不是我在英国遵守的那个规则，即：沿道路的左侧行驶。

为了完全说清楚这一点，让我再举一例。我的孩子们在奥克兰汽车驾驶学校学到这样一条停车规则：把方向盘打向右端，使你的车靠向路边，当你车的前轮与你前面车的后轮对齐以后，再把方向盘一直打回到左端。请注意，如果他们遵守了这条规

则，这条规则的意义就必然对他们行为的产生起着一种因果作用。听到这条规则，我感到很有趣，因为我并没有遵守这条规则。事实上，在我停靠车子时，我根本不遵守什么规则。我只是看准路边，留神不去碰撞前后车辆，尽可能地靠向路边罢了。但请注意，从表面或局外人来看，我的行为与一个遵守该规则的人的行为毫无二致。但我们还是不能够说，我遵守了这条规则。行为的形式特性不足以表明人们在遵守着一种规则。一种行为是否在遵守着一种规则，要看这种规则的意义是否在行为中起着某种因果作用。

现在我们可以把针对认知主义的这个讨论中的含义极简明地表述为：从人类遵守规则的意义上说(顺便提一句，人类遵守规则的情况远比认知主义所主张的要少得多)，计算机根本不会遵守规则。它们不过是按照一定形式程序去行动。计算机的程序决定着机器所要采取的各种步骤，决定着机器怎样从一种状态转变成另一种状态。我们可以打比方说，机器仿佛是在遵守规则。但从人类遵守规则的本意上讲，计算机不会遵守规则，它们只是在仿佛遵守规则的意义上行动。这种比喻的说法本无害处，的确，在科学上，这种比喻既普遍又实用。我们可以把任何系统都比作仿佛在遵守着规则，例如太阳系。只有这种比喻与其本来的含义相混淆时，才成为有害的。使用一种心理学的比喻去说明计算机是完全可以的。但是，一旦你把这种比喻当作本意来理解，一旦你使用计算机遵守规则的比喻去说明最初作为这个比喻基础的心理学意义上遵守规则现象时，混乱就产生了。

我们现在能够指出为认知主义提供的语言学证据错在何处了。如果说当人们谈话时，确实在遵守着语法规则的话，这决不能表明人们的这种行为与数字计算机是相同的，因为从人们遵

守语法规则的意义上说，计算机根本就不会遵守规则。它只能执行形式的程序。

这样，我们就有两种意义上的遵守规则，一种是在本来意义上的，一种是比喻意义上的。两种情况极易混淆在一起。现在，我想把这种区分应用到信息处理的概念上来。我相信，在信息处理的概念中包含着一个与此相似的重大混淆。这种混淆就是：因为我思考时是在处理信息，同样我的计算器在输入、传递、输出某些信息时，也是在处理信息，所以，我与计算器在处理信息过程中必定存在着某种一致性。在我看来，这显然是错误的。当我思考时我作信息处理，其含义是指我有意识或无意识地在参预着某些心理过程。而从这种信息处理的意义上说，计算器是作不了信息处理的，因为它根本不具有任何心理过程。计算器只是在模仿或模拟我心理过程的形式特点。也就是说，尽管计算器所经历的步骤与我所经历的步骤在形式上一模一样，也决不表明它的所作所为与我的完全等同，因为理由很简单，计算器不具有心理现象。在做6＋3时，它不懂得数字“6”代表着数目六，数字“3”代表着数目三，加号代表着加法运算。理由很简单，它对任何事物都是一无所知的。的确，也正是由于这个原因，我们才拥有了计算器。它们无需从事任何心灵的劳作，它们能比我们计算得更迅速更精确。它们无需经历我们所必须经历的信息处理过程。

因此，我们需要对这两种意义上的信息处理概念作一种区分，至少要把这两类根本不同的信息处理过程区分开来。第一类信息处理，我称为“心理学的信息处理”，它包含着心理状态。最笼统地说：当人们从事心理的操作时，实际上也就是在思想，从特性上说，思想包含着对种种信息的处理。另一种意义上的信息处理根本不具有心理状态。既然如此，这些处理过程只是

些仿佛进行某种心理信息处理的过程。让我们把这第二类信息处理称为信息处理的“仿佛”形式。假如我们不混淆这两类信息处理，使用这两种概念是毫无害处的。而我们所见到的认知主义正是坚持了这两者的混淆。

一旦我们看清了这种区别，我们就能够发现认知主义论据中的一个弱点，这也是它最根本的弱点之一。从我思考即是在进行信息处理的事实出发，从计算机进行信息处理——甚至是进行模拟我思考的形式特点的信息处理的事实出发，根本就无法得出任何在心理方面与计算机程序相关的东西。要想表明计算机程序与心理存在着关联，就必须有某个独立的论据，以说明这个“仿佛是”计算式的信息处理与心理相关。而这个信息处理的概念正被用来掩饰这种混淆，因为它是一种用来掩盖两种截然不同现象的表述。总之，我们在“遵守规则”这一概念中发现的混淆丝毫不差地存在于“信息处理”的概念中。

不过，在信息处理这个概念中，还存在着一个更隐蔽、更细微的混淆。请注意，从“仿佛是”信息处理的意义上说，我们可以把任何系统都看作仿佛是在进行信息处理。甚至，我们可以通过这种仿佛式的信息处理来收集信息。所以，并非只有使用计算器或计算机才可称得上是“仿佛式”的信息处理。例如，我们来考虑一下水顺山坡而下的情况。我们可以认为，水仿佛在进行信息处理。甚至我们还可以认为，我们从中得到了有关信息。比如说，通过它我们得到了有关山坡表层哪一面阻力最小的信息。但是，从以上事实可以看出，水顺山坡往下流并没有什么与心理相关的东西。关于对水的引力作用根本不存在什么心理学的内容。

我们可以把这一观点应用到对脑的研究上来。脑具有一个真实的心理的信息处理层次，这是一个显而易见的事实。换句

话说，人们实际思想着，思想存在于人脑中。而且，在脑神经生理层次上存在着各种各样的实际引起了我们思想过程的成分。但许多人设想除了这两种层次，即朴素心理学层次和神经生理学层次以外，一定还存在着某个额外的计算方面的信息处理层次。为什么他们会作此设想呢？在我看来，原因之一就在于，他们把心理方面的真实的信息处理层次与对脑过程进行“仿佛式”的信息处理描述的可能性混为一谈了。如果谈起水顺山坡而下的现象，没有人会认为这种现象与心理有什么关联。但是，用完全同样的观点去看待脑的问题，就较为困难一些了。

在心理方面与脑相关的事实是，脑容纳着心理过程，脑具有产生、体现这些心理过程的神经生理机制。但是，我们能够从一种“仿佛式”的信息处理观点去描述脑中其他过程这一事实本身，无法证明这些过程在心理方面是真实的，甚至与心理相关。一旦我们的话题深入到脑的内部，就不易发现这种混淆了。然而，这种混淆确实与我们以上所指出的那种混淆一模一样——因为水顺山坡而下进行了“仿佛式”的信息处理，就认为在顺坡而下的水中隐藏着某种心理。

接下来，我们要加以考察的是那种认为在所有富有意义的行为背后必定存在着某种内在理论的假设。人们不仅仅在认知心理学中，在许多领域中都可以发现这种假设。例如，乔姆斯基对一种普遍语法的研究所依据的正是这种假设。他假设，如果各种语言有着某些共同特征，如果这些特征又被人脑中的共同特性所制约，那么在人脑中就一定存在着一套极其复杂的普遍语法的规则。但对此还有一个更简单的假设，脑的生理结构，不经一种中间层次的规则或理论的介入，就决定了可能的语法。这个假设不仅更简单，而且也表明对那种由脑的固有特性所决定的语言的普遍特征，在神经生理的层次上加以描述就足够了。

人们根本就无需去假设，在神经生理的结构之上还存在着什么规则。

但愿我能用以下两种类比说清这个问题。我们人类的视觉不能够看见红外线和紫外线，这是一个简单的事实。难道这是由于我们有一种视觉语法的普遍规则，这种规则告诉我们："不要去看红外线或紫外线，"我们因此就看不见它们了吗？决不是这样，其所以如此显然是由于我们的视觉器官根本无法感知到光谱的这两端。当然，我们可以自认为我们仿佛在遵守着一种视觉语法的规则，然而事实上并非如此。再另举一例。如果我们试图对人行走时掌握平衡的能力作一种理论性的分析，其中似乎或多或少存在某种复杂的心理过程，我们好像由于领会了各种各样的提示而解决了一系列的二次方程似的，当然是在无意识的情况下，这些使得我们行走不至于跌跤。但实际上，我们知道没有必要用这种心理的理论去说明走路不跌跤的原因。事实上，人们这种掌握平衡的能力，在很大程度上是人们内耳中流体的功劳，而它根本不作任何计算。如果你一个劲儿地旋转，直到足以搅乱你内耳中的流体，那么你很可能会摔倒的。所以我想指出，我们的大量认知成果很可能是像这样获得的。脑恰恰是这样来提供认知成果的。我们没有充分的理由去假设除了我们心理状态的层次和我们神经生理的层次以外，还存在着某些无意识的计算过程。

我们再来考虑辨认人的面孔这种情况。我们都能毫不费力地辨认出我们的朋友、亲戚和熟人的面孔，确实，现已证明脑中的某些区域是专门用于辨认面孔的。为了弄清脑中的这种功能，假设我们设计出一台能够像我们人那样会辨认面孔的计算机。这台计算机做的完全是计算方面的工作，包括许多几何与拓扑方面的计算。难道这就证明了在我们辨认面孔的方法中也

包含有计算和运算吗？请想一想，当我们在湿沙地上踩出一个脚印时，我们的脚与沙地都不会做出任何计算。但如果我们要设计出一种程序，这种程序会根据作用在沙地上的不同压力所提供的信息来计算脚印的拓扑结构，这将是一项相当复杂的计算工作。但是，对一种自然现象进行一种计算方面的模拟包含着复杂的信息处理这一事实，决不表明自然现象自身也包含有这种处理。而事实很可能是，我们对面孔的辨认正像在沙地上踩脚印那样简单，那样自然。

实际上，如果对这种计算机的类比深究下去，我们就会发现计算机中的许多过程根本就不属于计算过程。例如，对某些计算器来说，如果你问："计算器是怎样做7乘以3的？"回答是："将3自加七次。"如你再问："计算器是怎样将3自加的？"对此就不存在任何计算方面的回答了，而这恰恰是在硬件中完成的。所以，对此问题的回答就是："计算器就是这样做的。"我想指出，对于我们大量的相当基本的能力来说，比如我们的视觉能力或学习一种语言的能力，在这些能力之下，不存在任何理论的心理层次。因为脑就是这样做的。正是因为我们具有这样的神经生理结构，感光器细胞上光子的作用就使我们能看见；也正是因为我们具有这样的神经生理结构，我们倾听别人谈话并与他们交谈的刺激就能使我们学习一种语言。

当然，我并不是说，规则对我们的行为一点作用也不起。相反，例如语言的规则或游戏的规则似乎就在有关的行为中起着决定的作用。但我认为，判定哪些行为受规则制约，哪些行为不受规则制约，这是一个棘手的问题。并且我们不能仅仅设想全部富有意义的行为都有某种规则的系统作为它们的基础。

也许在此更便于指出，尽管我对认知主义的全部研究工作并不抱有乐观的态度，但我确实认为，我们很可能能从这一研究

成果中获得许多有价值的见解。同时，我当然也不想阻止任何人去证明我的观点是错误的。即使我是正确的，我们仍能从失败的研究工作当中发现许许多多的洞见，行为主义与弗洛伊德的心理学就是两个恰当的例子。就认知主义来说，我特别看中大卫·马尔对视觉的研究和众人对“自然语言理解”的研究——设法使计算机去模拟人们日常言语产生和解释的研究。

在本章结束之际，我想以一种更明确的方式来表述为研究心灵而进行的这种探讨的内容。与认知主义所描绘的画面相反，让我们用另一种方法去解决困扰着社会科学的难题。我们要放弃那种认为在心脑之间存在着一种计算机程序的看法，还是把心与心理过程视作如同生长、消化、胆汁分泌一样基本的生物现象。例如，把我们视觉经验视作始于光子对视网膜的刺激，止于脑中某个部位的一系列活动的最终产物。那么，在对于动物视觉发生原因的解释上存在着两种总的描述层次。首先是神经生理学层次。在这个层次之上，我们对单个的神经元、突触以及动作电位进行探讨。但在这一神经生理学的层次之内，还有着较低和较高的描述层次之分。我们不必把自己仅仅限制在神经元和突触水平上。我们还可以去探讨神经细胞团或神经元模块的行为，比如像视网膜中各类神经细胞的不同层次或视觉皮层中细胞柱；并且我们还可以从更多的复杂层次上去探讨神经生理系统的性能；例如，在视觉中条纹皮质的作用，在视觉皮质中第18区和第19区的作用以及视觉皮质与脑的其他部位在处理视觉刺激时的关系。在神经生理学的层次之内，存在着一系列的描述层次，但所有这些都是属于神经生理学的。

除此之外，还存在着一个对人心的描述层次。例如，我们知道知觉是一种期待的功能。如果你期待着看到某物，你就会更

快地看到某物。此外我们还知道，知觉会受到多种心理现象的影响。一个人的心境与情绪会影响到他怎样感知以及感知到什么。同样，在人心的层次中也有不同的描述层次。我们不仅可以去探讨知觉是怎样受着个人的信念和欲望影响的，也可以去探讨知觉是怎样受着像个人的背景能力以及整个世界观这样的整体心理现象影响的，但是除了神经生理学的层次和意向性的层次以外，我们无需再去假设另一个层次，无需去假设一个数字计算处理的层次。只要我们不把信息处理的真实形式——心理的形式与"仿佛式"的信息处理形式混为一谈，把心理状态和神经生理这两种层次视为信息处理的层次就是毫无害处的。

那么最后，我们对认知主义的研究纲领应作如何评价呢？当然，我并没有证明它是错误的，也许它倒是正确的。但我认为，它成功的可能性与五十年前行为主义成功的可能性相比大概是差不多的。这就是说，我认为它成功的可能性事实上等于零。我之所以这样讲，所依据的只不过是以下三个方面：第一，我曾指出，一旦我们将认知主义背后的那些基本假设揭露出来，它们的荒谬性就是十分明显的。但是，这些假设在很大程度上已深深地嵌入到我们的理智文化中，其中一些极难加以根除，甚至有些已完全成为有意识的东西了。我所指出的第一点就是，一旦我们完全理解了这些假设的性质，它们的荒谬性就是显而易见的。我所指出的第二点是，我们确实没有足够的经验方面的证据去推测这些假设为真。因为对某些关键性的行为，像信息处理和遵守规则这些行为，现存的证据是模棱两可、可作多种解释的。第三，关于心脑之间的关系，不论是在本章还是在第一章，我提出了另一种观点。这种观点认为，我们不需要去假设，介于脑的神经生理性与心的意向性之间，存在着任何算法计算处理的中间层次。这幅画面，对于我们的讨论是非常重要的，其

特征就是，除心理状态的层次，诸如信念、欲望和神经生理的层次之外，不存在任何其他的层次，在心与脑之间不需要任何鸿沟的填充物，因为在二者之间根本就没有鸿沟可填。把计算机比作脑较之先前那些对脑的机械性比喻，大概既不高明也不逊色。我们知道，把脑说成是一台计算机同把脑说成是一部电话交换机、一套电报系统、一台水泵或一台内燃机是一样的。

假设没有一个人知道钟是怎么走的，假设要想弄清楚其中的奥秘又极为困难，因为，尽管到处都有钟，但已没人会制造，人们若要极力弄清钟的工作原理就会把钟拆毁了。现在，再假设有一组研究人员说："如果我们能设计出一种在功能上等同于钟，能像钟那样计时准确的机器，我们就能明白钟是如何工作的了。"于是，他们设计出一台沙漏并声称："现在我们明白钟是如何工作的了。"他们或许说："只要我们有了像钟一样计时准确的沙漏，我们就能最终搞清楚钟是如何工作的。"把这个比喻中的"钟"换成"脑"，把"沙漏"换成"数字计算机程序"，把计时的概念换成智能的概念，你就会了解人工智能与认知科学在当代的大致(但不是全部!)处境。

在整个考察中，我的全部目的就是试图去回答人类如何与周围宇宙相统一这些最令人困惑的问题。在第一章，我试图去解决"心身问题"。在第二章，我驳斥了把人类与数字计算机视为同一的某些极端的主张。在这一章，我向认知主义的研究纲领提出了质疑。在本书的后半部分，我想把注意力转到说明人类行动的结构、社会科学的性质以及意志自由的难题上去。

四 行动的结构

本章的目的是说明人类行动的结构。其所以要说明这个问题，有几个原因。我需要表明行动的性质如何与我前几章对心身问题的阐述和对人工智能的反驳相一致。我需要说明行动的心理部分并表明它与行动的物理部分是如何相关的。我要表明行动的结构如何与行动的说明相关。并且，我需要为讨论社会科学的性质和自由意志的可能性问题奠定一个基础，这两个问题我将在最后两章加以讨论。

如果我们考虑一下人的行动，我们立即会发现人的行动与自然界中的其他事件有某些显著的区别。起初，我们很容易把种种类型的行动或行为与种种类型的身体动作看成是一码事。这显然是错误的。例如，同一组人体动作可以构成一种舞蹈，可以表示在发信号、做体操、测试肌肉，或者均不表示这些内容。而且，正像同一组类型的身体动作能构成完全不同种类的行为一样，一种类型的行动也能通过大量的不同类型的身体动作来完成。例如，你想给一个朋友捎个信，你可以把它写在一张纸上，也可以用打字机打。你可以通过邮递员或拍电报来传送信息，还可以通过电话直接把信息告诉他。甚至，对于这些发送同一个信息的每一种方式，还可以采用各种各样的身体动作来完成。你可以用左手或右手来写，也可以用脚趾来写，甚至可以用你的牙叼着铅笔写。此外，行动还具有另一种使自己与事件一般地区别开来的独有的特征，即行动似乎具有优先描述的特征。

假如我要去海德公园散步，在我步行的过程中会伴有若干其他事情发生。但是对它们的描述并没有描述我有意向的行动，因为在行动中，我要去做什么在很大程度上取决于我想我要去做什么。例如，我也在向着大致巴塔哥尼亚的方向移动，我的头发颤动着，我的鞋子磨损着，我还推开了面前许许多多的空气分子。可是，这些额外的描述似乎没有一种涉及到我的这一行动的实质，也就是说，它们没有涉及到我行动之所以为我行动的东西。

与此相关，行动的第三个特征是，行为者非常明白他自己要去做什么。他不必把注意力转向自身，不必为选择哪一种行动而去作一番考察，至少不必试着去做。所以，如果你问我："你是想走到海德公园呢？还是想距巴塔哥尼亚更近些？"尽管我的身体动作可能会同时适合于回答这两个问题，但我还是会毫不犹豫地给你一个唯一的回答。

我们能够轻而易举地辨认和说明我们自己和他人的行为，这也是一个有关人类的显而易见的事实。我相信这种能力是建立在我们无意识地掌握着某套原则的基础之上的，这正像我们识别英语语句的能力是由于我们无意识地掌握着英语语法原则一样。我相信在我们表述以下这样一些普通常识性的事情时，存在着一套我们预先假设的原则。例如："巴兹尔投保守党的票是因为他认为保守党会解决通货膨胀问题"、"萨莉从伯明翰搬到伦敦是因为她认为在伦敦会有更多的就业机会，"甚至还包括对这样一些简单事情的表述："那边那个做着奇怪动作的人，原来是在磨一把斧子，或者在擦他的皮鞋。"

认识到存在这些理论性原则的人们，通常都仅仅把它们当作一种粗俗的理论加以嘲弄，并认为它们应当被某个更科学的对人行为的解释所替代。对这种主张我深表怀疑，这正同我对

另一种主张，即认为我们应抛弃英语语法隐理论，它是我们通过学习而获得的东西，深表怀疑一样。我对它们怀疑的理由是同样的：使用隐理论是从事行动的一个组成部分，同样，使用语法规则是讲话的一个组成部分。所以，尽管我们可以对它有所补充或去发现关于语言和行为的各种有益的其他东西，但是要想用某种对这种现象的外在的"科学"解释来取代这种隐含的并且是构成这种现象的组成部分的理论是根本不可能的。

如果亚里士多德和笛卡尔今天还活着的话，他们会完全理解我们对人类行为的大部分说明，只是他们理解不了我们对生物和物理现象的说明。通常用来说明这一点的理由是：一方面亚里士多德和笛卡尔掌握的是一种最初级的生物学和物理学的理论；另一方面他们所掌握的关于人类行为的理论也是最初级的，而当我们在生物学和物理学方面取得进展时，我们在对人类的行为说明方面却没有取得能与之相比的进展。我想提出另一种观点。我认为亚里士多德和笛卡尔同我们一样，已经拥有了一种关于人类行为的相当精致和复杂的理论。我还认为，许多带有假设性质的对人类行为的科学解释，比如像弗洛伊德的理论，事实上是使用了，而不是抛弃了，我们关于人类行为隐理论的原则。

到目前为止，我所说过的几点归纳起来就是：种种类型的行动较之种种类型的身体动作还具有一些另外的内容，行动具有优先描述的特性，人们不加注意就知道自己要去做什么，以及我们用来识别和说明行动的原则就是行动自身的一部分，也即是行动的组成部分。现在我想对我们所谓行为的结构作一个简洁的描述。

在说明人的行为的结构之前，我先介绍一两个技术性的术语。在行为的结构中，关键性的概念是意向性的概念。说一种

心理状态具有意向性，意思不过就是说这种心理状态涉及到了某物。例如，一种信念总是一种这样或那样事情的信念，一种愿望总是一种这样或那样可能发生或可能存在事情的愿望。意向，从普通意义上说，在意向性理论中并不占有特殊的地位。意欲去做某事同有信念、有欲望、有期待、有恐惧等等一样，只是意向性的一种。

一种意向状态，比如一种信念、一种愿望或一种普通意义上的意向，从其特性上来看，有两个组成部分。一个是我们可以称为"内容"的部分，它使心灵涉及到某物，另一个是它"心理的方式"或"类型"的部分。我们之所以要作这种区分，是因为在不同的类型中可能存在着相同的内容。例如，我想要离开这个房间，我相信我将要离开这个房间，我有要离开这个房间的意向。在以上每一种情况中，都具有同样的内容，即我将要离开这个房间；但这一内容分别存在于不同的心理方式或类型中，即分别存在于信念、愿望和意向中。

再者，意向状态的内容与类型有助于将心理状态与世界联在一起，而这种联系正是我们的心具有心理状态的原因：让世界在我们面前再现；描绘它是怎样一副样子，怎样是我们所喜欢的样子，怎样是我们担心可能变成的样子以及对于它我们打算做些什么等等。由此就会出现这样的结果，如果我们的信念与世界上的情形相符，我们的信念即为真，如果不符则为假；同样，我们的愿望或者实现或者落空，我们的意向或得到执行或者得不到执行。因此，一般说来，意向状态具有"满足条件"。每一种状态在它为真的条件下（如果这种状态指的是一种信念），或在它实现的条件下（如果这种状态指的是一种愿望），或在它得到执行的条件下（如果这种状态指的是一种意向）确定自身。在每种情况中，心理状态都有它自己的满足条件。

关于意向状态，我们要注意的第三个特征是，它们有时导致事情发生。例如，我想去看电影，并且我确实去看电影了，这就是说在正常情况下，我的愿望将导致它所表现出的那个事件，即我去看电影。在这种情况下，原因与结果之间存在着一种内在的联系，因为这种原因是对它所导致的那种事态的一种表现。这种原因既表现又引起了这种结果。我把这类因果关系称为"意向的因果关系"状况。我们将会看到，对于人的行动的结构和说明，意向的因果关系将具有决定性的意义。这种因果关系在种种方面与标准教科书上所描述的因果关系大不一样，在教科书中，因果关系是像一个台球碰撞另一个台球并引起它移动的那种状况。对我们来说，意向的因果关系中基本的东西就是我们将要考察的那种状况，即人心导致它一直在思考的那种事态的状况。

将以上对意向性的讨论归结起来，在我们对人行为的分析中，我们需要记住意向状态的三个特征：第一，意向状态由存在于某种心理类型中的一种内容构成。第二，它们取决于它们的满足条件，也就是说，意向状态是否满足取决于世界与意向状态的内容是否相符。第三，它们有时导致事情发生，它们通过意向的因果关系导致一种符合，即导致它们所表现的事态，导致它们自己的满足条件。

现在，我要运用这些思想回过头来讨论本章的主要问题。我曾说过，要对可被称作是行动结构或行为结构的问题，作出一个非常简洁的描述。在这里，我所说的行为是指人的自愿的、有意向的行为，它是指像走路、跑步、吃饭、作爱、选举、结婚、买卖、度假和工作这样一类事情，而不是指像消化、长大或打鼾这样一类事情。但是即使我们把自己限制在意向行为的范围内，我们面前的人类活动仍然呈现出种种令人眼花缭乱的类型。我们需

要区分个体行为与社会行为；需要区分集体的社会行为与一种社会集体中的个体行为；需要区分出于其他目的的行为与以行为本身为目的的行为。也许最困难的就是，我们需要随着时间的推移去描述行为旋律的序列。人类活动毕竟不像一系列普通摄影的快照，而更有些像我们生活的电影。

我不打算去回答所有这些问题，但我确实希望我所要表述的东西最终会成为一种关于行动结构的常识性的描述。如果我是正确的，我所表述的东西就应该具有显而易见的正确性。但是，从历史上看，我所认为的这种常识性的观点好像并不是显而易见的。原因之一是，哲学和心理学中的行为主义传统使许多人忽略了行动的心理部分。行为主义者想利用纯粹的身体动作来定义行动，甚至我们的全部精神生活。有些人曾把行为主义的态度形容为"佯装麻醉状"，在我看来这种形容不是没有根据的。在哲学中另一个相反的极端认为，我们在任何时候完成的唯一行为就是内部心灵的意志行为。按照这种观点，严格地说，我们任何时候都举不起我们的手臂。我们做的一切只是举起我们手臂的"意志"。如果手臂确实举起来了，那完全是侥幸，但确实不是我们的行动。

另一个困难是，直到最近，行动哲学仍是一个有些受到人们忽视的论题。西方的传统始终强调知比行更重要。知识和意义理论在它们所涉及的领域中一直处于比行动理论更重要的地位。现在，我力图把行动的心理方面与物理方面放在一起加以研究。

对行为结构的说明最好能用一组原则加以表述。这些原则应该说明行动的心理和物理两个方面。在表述这些原则时，我将不讨论我们的信念、愿望等等是从哪里来的，但我将说明它们在我们行为中是怎样表现的。

我想，让大家理解这些原则的最简便的方式就是先把它们陈述出来，然后再试图进行辩护。请看以下原则。

原则1：行动，就其特性而言由两部分组成，即由心理部分和物理部分组成。

例如，你设想一下推一辆小轿车时的情景。一方面，在你推车时，你有某种有意识用力的经验。如果你成功了，那么这些经验就导致你身体的运动和小轿车相应的运动。如果你没有成功，你至少仍具有心理的部分，也就是说，你会有一种至少用某些物理的部分试图推动小轿车的经验。你的肌肉还处于紧张状态，你还有紧抵车子的感觉等等。由此得到

原则2：心理部分是一种意向。它具有意向性，即它涉及到某物。它确定着行动中所谓成功和失败；在成功的情况下，它导致了身体运动，身体运动又导致了其他的运动，比如像小轿车的运动。按照我们刚刚概述过的意向性理论，行动由两部分组成，即由心理部分和物理部分组成。在成功的情况下，心理部分导致物理部分并且表现物理部分。我把这种因果性形式称为"意向的因果关系"。

要想认清一种行动的这两个不同部分的性质，最好的方法就是把每个部分分离出来，分别地加以考察。事实上，这在试验室中是很容易做到的。在由蒙特利尔的怀尔德·彭菲尔德做的神经生理学试验中，我们已经看到了这种方法。彭菲尔德用电刺激病人运动皮质的某个部位，用这种方法能导致该病人肢体的运动。对此，病人们总感到吃惊，他们特别喜欢说这类话："我没有动，是你动的。"在此情况下，我们已将没有意向的身体动作分离了出来。请注意，这种情况下的身体动作可能与一种有意向的行动中的身体动作是完全一样的，但似乎又显然存在着一种区别。这种区别是什么呢？好！我们再来看看很早以前威

廉·詹姆斯的实验，通过这些实验，我们能够把心理部分分离出来，而不带有与此相应的行动的物理部分。按照詹姆斯的实验，先将病人的一只胳膊麻醉，在暗室内将他的这只胳膊握住，然后命令他举起这只胳膊。他照自己认为是服从命令的那样去做了，但是却十分吃惊地发现，他的这只胳膊根本就没有举起来。在这种情况下，我们将心理部分，即意向，从身体运动中分离了出来。因为这个人确实具有意向，也就是说，我们可以肯定他的的确确想要移动他的胳膊。

在正常情况下，这两个部分是结合在一起的。通常我们既具有意向又具有身体的动作，但两者不是彼此独立的。我们首先给出的这两条原则力图明确表达这两部分是如何相关的。心理部分作为满足条件的部分，必须既表现又导致物理的部分。顺便还请大家注意，要描述这些原则的内容，我们还有相当一批有关“试图”、“成功”、“失败”、“有意向”和“无意向”以及“行动”和“动作”等方面的词汇。

原则3：那种对行动的结构和行动的说明都是必不可少的因果关系就是意向的因果关系。我们行动中的身体动作是由我们的意向引起的。意向使事情发生，所以它们是原因；但它们也具有内容，因此能在逻辑推理过程中出现。它们可以既成为原因又具有逻辑特征，因为我们所说的这种因果关系是心理的因果关系或意向的因果关系。在意向的因果关系中，心理的内容影响着世界。整个心理机能的运转是在脑中实现的，正像我在第一章里所说明的那样。

我们在这里讨论的这种因果关系形式与哲学教科书里所描述的那种标准的因果关系形式有着相当大的区别。它不是一个规则性或涉及到法则和永久关联的问题。事实上，我认为它非常接近于我们常识性的因果关系概念。我们常识性的因果关系

概念不过是指某物引起它物发生的情况。而意向的因果关系其特殊之处就在于，它是一种心理状态引起它物发生的情况，并且，那种它物就是由导致它的心理状态所表现出的那种事态。

原则4：按照行动的理论，那些预先思考的、作为某种预先计划结果的行动与那些本能的、不加任何预先思考的行动，存在着一种根本的区别。与这种区分相对应，我们还需区分在先的意向与行动中的意向，前者是指在从事一个行动之前形成的意向，后者是指在我们实际从事一个行动过程中的意向。

在行动的理论中，一个常见的错误是，假设全部有意向的行动都是某种深思熟虑的结果，即一系列事实推理的产物。很显然，我们在做许多事情时，情况并非如此。我们做事完全可以没有任何在先的反思。例如，在通常的会话中，人们不去反思他们往下所要说的东西，他们只需说就是了。在这种情况下，确实存在着一种意向，但它不是在从事行动之前就形成的一种意向。它就是我所说的行动中的意向。可是，在其他情况下，我们确实具有在先的意向。我们反思我们想要什么以及什么是获得它的最佳方法。这种反思的过程（亚里士多德称它为"事实推理"）就其特性而言，或者导致一种在先意向的形成，或者像亚里士多德也曾指出的那样，有时引起行动自身。

原则5：在先意向的形成，一般说来，至少是事实推理的结果。事实推理总在推理如何在冲突的愿望中确定最佳抉择。大多数人（和动作）行动背后的动力是愿望。而信念，它的作用只在于能使我们断定怎样才能最佳地去满足我们的愿望。例如，我想去巴黎，并且我相信在所有考虑到的方案中，坐飞机去是最佳的方案，于是我形成了坐飞机的意向。这就是一个典型的和常识性的事实推理的例子。但是，事实推理在根本上与理论推理不同，与通过推理确定对象不同，在事实推理中，总是涉及到

如何在我们种种冲突的愿望中择其最佳决断。例如，假如我确实想去巴黎,并且我相信最好的办法就是坐飞机去。然而,如果我不打消我大量的其他愿望,我就无法去坐飞机。我不想破费;我不想在机场排一个个的队;我不想坐飞机上的座位;我不想吃飞机上的食品;我不想让别人把胳膊肘放在我放胳膊肘的地方以及种种其他愿望。尽管只要我坐飞机去巴黎，所有这些愿望就会被放弃，我仍会从所有考虑到的愿望中推理出坐飞机去巴黎是最佳的抉择。这不仅是典型的事实推理，而且我认为这种涉及到对各种冲突愿望进行裁决的事实推理在事实推理中是普遍存在的。

于是，以上这五条原则在我们面前呈现出这样的画面：赋予行动以动力的心理之能是一种按意向的因果关系起作用的能量。它是能量的一种形式,按照意向的因果关系,或以愿望或以意向的形式,表现它所导致的那个事态。

现在让我们回过头来看看我们一开始提到的关于行动的那几个观点，因为我觉得我们已经有了足以说明它们的原则。我们注意到行动具有优先描述的特性,并且事实上,常识能使我们去识别什么是行动的优先描述。现在，我们知道了一个行动的优先描述是由行动中的意向决定的。一个人实际在做什么，或至少他试图做什么，完全是一个在他的行动中带有什么意向的问题。例如,我知道我要去海德公园而不是去巴塔哥尼亚,因为这是我去散步的意向。我不加注意就明白这一点，因为上述知识不是我外部行为的知识,而是我内部心理状态的知识。

而且这表明了我们对人行动说明的某些逻辑特征。说明一个行动就是去说明它的原因。它的原因就是心理状态。这些状态或者通过导致意向的事实推理步骤或者通过意向自身涉及到行动。对行动说明的这个最重要的特点仍然值得用另外一个原

则加以陈述。让我们把它叫作

原则6：对一个行动的说明必须与行为者在行动时或在他作行动意向的推理时头脑中的内容相同。如果这种说明确实是说明的，那种通过意向的因果关系方式引起行为的内容就必定与这种行为说明的内容同一。

从行动不同于世界中其他自然事件这一点来看，对它们的说明也就相应地有所不同。当我们说明一次地震或一次飓风时，说明的内容只需描绘发生了什么和为什么发生就够了。它不会实际导致事件自身。但是在说明人的行为时，原因和说明都具有内容，并且这种说明所以成其为说明是因为它具有与原因同样的内容。

迄今我们所讨论的仿佛是，人们仅仅是在一瞬间就具有了意向。这当然与现实是极不相符的。现在我们有必要介绍一些复杂的情况，这至少会使我们的分析更接近现实生活中的事情。没有人从来就只有一种意向，一种自身孤零零的意向。例如，我有一个从伦敦驱车前往牛津的意向；我的这种意向可以带有相当的自发性，但我一定仍然具有一系列其他有意向的状态。我一定具有一个我拥有一辆小轿车的信念和一个牛津在我的行程之内的信念。而且，从特性上看，我还会有一个道路不要太拥挤的愿望和一个天气不要太坏以免不利于行车的希望。从特性上看，我不仅仅是驱车牛津，还是出于某种目的去的（这里有点接近于行动说明的概念）。照此说来，在特性上，我从事的是事实推理，即那种不导致信念或论证结论而导致意向和实际行为的推理形式。我们理解了这种推理形式，就会向理解行动的说明迈进一大步。让我们把给我的意向状态以特殊意义的那些意向状态称为其他的意向状态，把所有这些意向状态称为“意向性之网”。我们用一个一般性的结论来表述，把它称作

原则7：任何意向状态只是作为一个其他意向状态之网的一个部分起作用的，这里所说的"起……作用"，是指这个部分相对所有其他意向状态来说仅仅确定了它的满足条件。

在我们开始仔细考察这张意向之网细节的时候，我们又发现还有另一个有趣的现象。这就是我们的心所从事的活动不存在于心理状态之中，甚至可以完全肯定这一点。更确切地说，我们的心理状态只是以它们自己的方式起作用的，因为非意向状态的能力、才能、技艺、习惯、做事的方式以及对待世界的一般态度，只是作为它们起作用的背景而存在的。对我来说，想形成开车去牛津的意向，我必须具有开车的能力。但是这种能力自身不存在于整个众多的其他意向状态之中。要想把车开走，就要具备比一串信念和愿望更多的东西。也就是说，我必须确实具备开车的技术。这就是一个有关我知道怎么一回事不完全等于知道怎样去做的问题[①]。我们把这个技艺、习惯、才能等等的集合，即意向状态起作用所依靠的集合称作"意向性背景"。对于意向性之网的那个观点，即任何意向状态只是作为一种意向性之网的一个部分起作用的观点，我们还要加上这个意向性背景的观点。我们把它称作

原则8：整个意向性之网是以不属于心理状态的人的能力为背景起作用的。

我曾说过，许多被想象成为科学的行为观试图逃离或超越我所描绘的这种常识性模式。但我认为它们终将无路可走，因为这些对行为描述的原则恰恰没有描述这样的现象；它们自身就是构成这些现象的组成部分。例如，看看弗洛伊德的说明。

---

① 英国哲学家吉尔伯特·赖尔（1900—1976年）在他的《知道怎样与知道那样》（1946年）一文中最先提出这种区分。——译者

当弗洛伊德炮制他的心理玄学的时候，也就是当他为他的工作提供一种理论的时候，他常常使用科学上的比较。他认为，在心理学与电磁学或水力学之间存在着许多相似之处，我们要按照液压原则来类推心理的功能，等等。但是，当他实际去检查一个病人时，当他实际去描述某个病人的神经症性质时，令人吃惊的是他所提出的说明许多都是常识性的说明。杜拉的行为之所以那样，因为她在与K先生恋爱，或因为她在模仿那个已经去玛丽亚采尔的堂姐[1]。弗洛伊德的这个观察结论对常识有所增加之处在于，导致我们行为的心理状态常常是无意识的。的确，它们是受到压抑的。我们常常拒绝承认我们的某些意向状态，因为我们为承认它们而感到羞愧，或因为某些其他的原因。其次，他还贡献了一种心理状态转换的理论，即关于一种意向状态怎样被转换成另一种意向状态的理论。但除了这些和其他这类观点以外，弗洛伊德的说明形式与常识的形式完全一样。我认为，恰恰是在我们获得其他更科学的行为观的时候，这种常识会继续保留下来。既然说明的结构必须与被说明现象的结构相符合，对说明的改进就不可能具有新的和闻所未闻的结构。

在这章，我试图说明的是行为怎样以及在什么意义上既包含着心理状态又由内在的心理状态引起。许多心理学和认知科学试图否认这些关系，这似乎是令人吃惊的。在下一章，我将深入研究我的这个关于人行为的观点应用在社会科学上产生的某些结果。为什么说社会科学遭到了失败同时又取得了它所应有的成功？我们能明智地期待着从社会科学中学到些什么？

---

[1] 见弗洛伊德《少女杜拉的故事》。——译者

五　社会科学的前景

在这章，我所要讨论的是当代最令人困惑不解的理智难题之一：为什么自然科学方法在人类行为的研究领域中没有给我们以像在物理学和化学领域中的那种回报？不管怎么说，我们能明智地期待着一种什么样的“社会”或“行为”科学？我将指出在人的行为与自然科学所研究的现象之间存在着某些基本的区别。我将辩明这些区别既解释了我们在人文科学中的失败，又解释了其成功之处。

首先，我想提醒你们注意一个重要的区别，即对人行为常识性的说明形式与标准的科学说明形式之间的区别。按照科学说明的标准理论，说明一种现象就在于表明这种现象是怎样遵循着某些科学法则发生的。这些法则是关于事物怎样发生的普遍概括。例如，如果你得到了一种描述一个下落物体行为有关法则的陈述，同时你还知道下落物体的起点，那么你就确实能推演出这个下落物体将会发生什么。与此相似，如果你想要说明一条法则，你可以从某个更高层次的法则来推演出这条法则。因此，说明与预言是完全对称的。你通过推演去预言将会发生什么；你通过推演去说明已经发生了什么。不管这种类型的说明在自然科学中会具有什么样的价值，在这章，我所强调的观点之一就是，这种类型的说明对于我们说明人的行为是毫无价值的。这倒不是因为我们缺乏去说明人行为的个体实例的法则，因为即使我们有了这样的法则，仍然是无济于事的。如果我们当真

有了这样一种“法则”，即一种涉及到你行为某个方面的普遍概括，我想我会轻易地让你想象到它会是一种什么样的法则。

假设在最后的选择中，你投了保守党的票，假设你之所以要投票赞成保守党是因为你认为保守党会比其他党派更能够去解决通货膨胀问题，假设你为什么投保守党的票这个事实与你投了保守党的票这个事实一样都是一种很普通的事实，再假设某些政治社会学家针对与你的状况——社会经济地位、收入水平、受教育程度、其他利益等等——完全相符的人们提出一种绝对没有例外的普遍概括。假设这种绝无例外的概括其意思是说与你的状况相同的这类人一定会去投保守党的票，那么现在我要问：哪一种说明能回答你为什么投保守党的票这个问题？是用你诚心诚意地赞成它的那个原因来说明？还是用那个普遍的概括来说明？我要加以辩明的是我们决不会接受用那种概括作为对我们自己行为的说明。那种概括陈述了一种规律性。这样一种规律性的知识可能对于预言是有用的，但它说明不了关于人行为的任何个体情况。对此确实需要进一步地说明。例如，为什么所有那批人都投保守党的票？一种答案表明，你投保守党的票是因为你担心通货膨胀——也许你们这批人都特别为通货膨胀所左右，而这正是他们都以同样的方式参加选举的原因。

总之，我们不同意用一种概括来说明我们自己或任何其他人的行为。如果发现了一种概括，这种概括本身首先就需要我们所追加的那种说明。就人的行为而论，我们通常所寻求的是这样一种说明：具体地去说明心理状态，即信念、恐惧、希望、愿望等等，在人行为的产生中所起的原因作用。这是我在上一章所描述过的。

让我们再回到我们最初的问题上来吧，为什么我们似乎不具有在我们具有自然科学法则意义上的那种社会科学的法则？

对这个问题有几种一般性的回答。有些哲学家指出，我们不具有一门关于行为的科学与我们不具有一门关于家具的科学原因是一样的。我们之所以不可能具有这样一门科学是因为不存在椅子、桌子、写字台以及所有其他种类的家具所共有的物理特性，以至不能使它们归属到一套共同的家具法则之下。除此以外，我们确实也不需要这样一门科学，因为我们要加以说明的东西——例如，为什么木头桌子很坚固？为什么铁网家具会生锈——已经能被现今的科学加以说明了。同样，全人类的行为没有任何共同的特征。此外，我们希望得到说明的人行为的特殊方面能够由物理学、生理学以及当今所有的其他科学去说明。

在一个与此有关的论证中，有些哲学家指出，也许我们用来描述我们自己和其余人类的概念根本就不会恰好符合像物理学和化学这样的基础科学的概念。他们指出，也许人文科学好像一门关于天气的科学。我们拥有一门关于天气的科学，即气象学，但它不是一门严格的科学。因为我们所关心的气象问题与我们用于物理学的自然范畴不相符合，像"英格兰中部地区上空有光点"或"伦敦多云"这样的气象概念与物理学的概念分属两个系统。在杰里·福多尔的著述中，我们可以看到对这种观点的一种有力的表述。他指出，像地质学或气象学这类专门科学所研究的是以种种方式体现在物理学中的世界特征，并且专门科学与更基础的物理科学之间的这种松散的联系也是社会科学的特性。正像山脉和暴风雨是在不同种类的微观物理学结构中实现的一样，货币也能在物理属性上体现为金、银和纸币。像这样较高层次的现象与较低层次的现象之间分离性的联系，的确允许我们拥有一大批科学学科，但是这些学科不考虑严格的法则，因为这种松散的联系形式将允许法则出现例外。

对这种观点，即我们不可能具有联结心理与物理的严格法

则的观点，还有另一种论证，这是由唐纳德·戴维森提出来的。他认为，合理性、一致性以及连贯性的思想在一定程度上构成了我们心理现象的概念，而这些概念没有系统地与物理学的概念发生关联。用戴维森的话来说，这些概念在物理学中没有"共鸣"。不过，这种观点所面临的一个困难是：有许多科学，它所包含的基本概念同样在物理学中没有"共鸣"，但它们却依然属于相当坚实的科学。例如，生物学离不开有机体概念，而"有机体"在物理学中是没有"共鸣"的，但是这并没有妨碍生物学成为一门硬科学。

另一种相当普遍的意见认为，我们心理状态复杂的相互关系永远妨碍着我们得到一套将它们与神经生理状态联结起来的系统法则。按照这种观点，心理状态卷入了复杂的相互关系之网，因此不能系统地与种种脑状态相对应。但是，这种论证也是缺乏根据的。例如，假设诺姆·乔姆斯基下述思想，即我们每个人生来在脑中就被编入一套复杂普遍语法规则的思想，是正确的，那么普遍语法规则的复杂性和相互依赖性决不会妨碍普遍语法规则系统地体现在脑神经生理学中。相互依赖性和复杂性本身不是否定严格心理-物理法则可能存在的一个充分证据。

我发觉所有这些观点都是富有启发性的，但我不认为这些观点准确地抓住了心理与物理科学之间真正基本的区别。一方为社会学和经济学，另一方为物理学，双方之间的关系的确十分不同于像气象学、地质学、生物学以及其他具体自然科学与物理学的关系。我们有必要努力揭示其中的原因。理想地说来，我希望能够为你们提供一种一步步的论证，以表明严格的社会科学之可能性的限度以及这些学科的真实性质和力量所在。我认为我们必须彻底抛弃那种把当今的社会科学视作牛顿之前的物

理学的观念，必须彻底放弃那种等待着一套牛顿式的心灵与社会法则的幻想。

首先，这个假设的难题究竟是什么？人们会说："社会和心理现象无疑与其他现象一样真实，那么为什么没有关于社会和心理行为的法则呢？"为什么应该存在分子行为的法则，而不存在社会行为的法则？否证一个论点的方法是先把它设想为真，然后去说明该假设由于某种原因而成为谬误。让我们先假定我们实际具有了社会法则和历史法则，这些法则能使我们对战争的爆发和革命的来临有所预见。我们还假设，我们对战争和革命的预见其准确和精密的程度如同我们在海平面的一个真空中预测一个下落物体的加速度一样。

这个真正的难题就是：不管什么样的战争和革命都包含着许许多多分子运动，但在这里会有一个这样的推论：任何关于战争和革命的严格法则应该完全与关于分子运动的法则相符合。要使一场革命在某一天爆发，有关的分子就必须按一定的方向移动。如果是这样，在革命及其参加者层次上根据那种预言革命的法则所作出的预言就将与在物理粒子层次上根据分子运动的法则所作出的预言相同。现在我们可以把最初的问题转变成这样的问题：为什么那种较高层次即革命层次的法则不能与那种较低层次即粒子层次上的法则完全符合呢？好，要看清这两者为什么不能符合，先让我们考察一下在较高和较低层次的法则之间真实存在着的某些完全符合的例子，然后我们就能认清这些情况与社会的情况存在着哪些不同。

将一层次上的法则化约为较它更低层次上的法则，其中最成功的例子之一就是气体定律即波义耳定律和查理定律向统计力学定律的化约。这种化约是如何进行的呢？气体定律讲的是压力、温度、气体体积之间的关系。例如，气体定律预言，如果将

装在圆柱形容器中的气体加温，就会增加气体对该容器壁的压力。统计力学定律涉及到微小粒子群的行为。例如，它预言如果增加了气体中粒子的运动速率，就会有更多的粒子更猛烈地撞到容器壁上。我们之所以能在这两套定律之间找到一种完全的符合，原因就在于我们能够完全利用粒子的活动去说明温度、压力和体积。增加了气体的温度也就增加了粒子的速度，增加了撞击容器的粒子数量和速度也就增加了压力。由此得出，增加温度将会导致压力增加。为了讨论起见，假设情况不是像这样，假设根本就不存在用更基本的粒子活动来说明压力和温度的情况，那么压力和温度层次上的任何定律都将是不可思议的。因为压力和温度存在的方式恰恰与粒子存在的方式在时间和空间上是同一的，如果压力和温度层次上的系统行为与粒子层次上的系统行为之间不存在系统的关系，那将是不可思议的。

这是一个很简单的例子。让我们再举一个稍复杂一点的例子。摄取热量要与消耗热量保持平衡，否则会增加或者减少脂肪储存，这是一个"营养科学"的定律。或许这不是一个花哨的定律，而是一个相当现实的定律。它具有为我们大家所熟悉的推论：如果你吃得多又不进行足够的锻炼，就会发胖。与气体定律不同，这个定律不是以简单的粒子活动方式为依据的。要想为这个定律提供粒子活动的依据不是一件简单的事，因为，比如说在活的机体中，食物被转换成脂肪储存要经过一系列相当复杂的过程。然而尽管复杂，这个定律仍然有一个由更基本的粒子活动提供的依据。在其他条件不变的情况下，一旦你吃多了，分子就会精确地朝着一定的方向汇集，使你发胖起来。

现在我们可以进一步朝着我们的结论推进。从存在着气体和营养定律这一点上说，将不会存在战争和革命的法则。我们使用像脂肪储存和压力这样的概念来把握的那些现象是系统地

以更基本层次上元素活动为依据的，在这一点上，我们使用像战争与革命、婚姻、货币与财产这样的概念来把握的那些世间的现象却不是系统地以更基本层次上元素活动为依据的。请注意，正是有了这种依据才表明我们在一门科学的较高层次上取得了重大进展。DNA结构的发现之所以对生物学这么重要，疾病的病菌理论之所以对医学这么重要，原因就在于两者都从更基本的元素出发提出了系统说明像遗传特性和病症这样较高层次特征的承诺。

但是这样一来新的问题又产生了：如果说社会和心理现象没有按这种方式找到依据，那么为什么没有？为什么不能找到？退一步说，就算它们没有这样的依据，那么为什么没有？像一切其他事物一样，战争和革命是由分子运动构成的。那么为什么像战争和革命这样的社会现象不能像热量摄取和脂肪储存之间的系统关系那样系统地与分子运动发生关联呢？

要搞清社会现象为什么不能如此，我们不得不问，社会现象都有哪些能使我们把它们列为范畴的特征？在我们用范畴把握心理和社会现象时遵循哪些基本原则？一种极为重要的特征是：对大量的社会和心理现象来说，指称一现象的概念自身就是这一现象的一个组成部分。要想把一事物视作婚礼、工会、财产、货币甚至战争与革命，卷入到这些活动中的人们就必须具有某些相应的思想。一般说来，人们不得不对这些活动是什么有所思考。例如，为了结婚或购买财物，你和其他人就不得不对什么是你们所要做的事情有所思考。这就是社会现象的一个至关重要的特征。而在生物和物理科学中就不存在这种情况。即使没有人去想："这是一棵树，或一种作物或一例结核病"，某物仍是一棵树或一种作物，某人仍患有结核病。即使根本没有人去想它们，它们依然存在。但是许多描述社会现象的词语则不得不成

为社会现象的组成部分。从中我们进一步得出：这样的词具有一种特殊的自我涉及。“货币”涉及到凡是人们当作货币来使用和认作货币的事物；“允诺”涉及到凡是人们有意作出的允诺或视作允诺的东西。我的意思不是说人们要建立货币制度就必须具有那个词或在他们的词汇中与那个词完全同义的某个词。而我所要表达的意思是：人们必须具有旨在把某物当作货币的某些思想和态度，并且这些思想和态度就是那个货币定义的一部分。

从这个特征还可得出一个极为重要的推论。对这样的社会现象加以规定的原则没有为被当作是社会现象的物理承担者的东西设立任何物理的限制。这就意味着在现象的物理与社会或心理特性之间不可能存在任何系统的关联。我们所说的社会的特征，部分地是由我们对它们所采取的态度决定的。我们对待社会现象的态度是不受我们所说的现象的物理特性制约的。因此，就不可能存在任何使严格的社会科学法则成为可能所必需的那种心理层次与物理层次间的符合。

在社会科学和自然科学之间存在着一种根本性的中断，对此进行论证关键性的一步所依据的是社会现象的心理特性。我们在前面介绍的那些类比，即用气象学、生物学和地质学对社会科学作的类比，正是忽略了这个特性。以社会和心理的学科为一方，自然科学为另一方，双方之间的根本中断根源于社会和心理学科中心灵的作用。

再来看福多尔的主张，他认为社会法则有例外是由于社会层次的现象松散地或分离地覆盖在物理现象之上。这种观点也没有揭示出我一直强调的那种根本中断。即使这种分离已在某一点上得以调整，但对另一个人来说总可以以无限多的方式来增加这种分离。假设货币一直采用的是一批有限的物理形式，例如金、银和纸币。但对另一个人或另一个社会来说仍可以把其

他东西作为货币。确实，只要作为物理承担者的那种材料能够被当作一种交换的媒介来使用，物理承担者就与货币的属性无关。

有人可能会这样反驳："那么好，为了获得精确的社会科学，我们不需要在世间的事物特性之间寻找一种严格的符合，我们只需在心理特性和脑特征之间寻找一种严格的符合。经济学和社会学在物理世界中的真正依据不存在于我们周围中所见的那些客体属性中，而存在于脑的物理属性中。所以，纵然认某物为货币的思想对于某物成其为货币是必不可少的，这种规定货币的思想仍很可能是，而且在你看来更是，脑中的一个过程。所以，要想表明不可能存在任何严格的社会科学法则，你就必须要表明在种种心理状态和种种脑状态之间不可能存在任何严格的关联，而你却没有表明这一点。"

为搞清楚为什么不可能存在这样的法则，让我们来考察我们似乎有可能得到的一门严格的神经生理学的若干方面，即关联着心理现象和神经生理现象的严格法则的若干方面。考虑一下痛觉。设想痛觉的神经生理原因至少对人类来说是有限的和确定的，这似乎是合情合理的。在前面的一章里我们确实对其中的一些原因进行过讨论。看来要获得一门完整的痛觉神经生理学在原则上不会有什么困难。那么视觉的情况如何呢？同样，要得到一门成熟的视觉神经生理学在原则上似乎也没有什么障碍，我们甚至可以对某些种类的视觉经验的神经生理状况作详尽的描述，例如，对我们看见红色东西的经验。在我看来，没有什么能妨碍我们去获得这样一门神经生理的心理学。

但是现在困难的地方在于：尽管我们可以在神经生理学和痛觉之间或在神经生理学和有关红色的视觉经验之间找到系统关联的关系，但我们不能对视某物为货币的神经生理学作出与

此相似的解释。为什么不能呢？假定每当你看见你面前的一些货币就会有某个神经生理过程产生，那么是什么妨碍了神经生理过程每次都以同样的过程出现呢？从货币具有无数种物理形式这一事实来看，货币对我们的神经系统能够产生无数种刺激效果。由于货币对我们的视觉系统具有无数种刺激样式，如果这些刺激样式对脑能产生一模一样的神经生理效果，那将又会产生一个奇迹。

这种适用于视某物为货币的情况甚至更适用于相信某物为货币的情况。如果在任何语言和文化中，一个人感到自己缺钱花的信念每次都具有完全相同类型的神经生理的现实过程，那简直是不可思议的。说到底是因为能产生那种信念的神经生理刺激其可能的范围是无限的。与神经生理学的情况相反，心理影响物理的方式始终妨碍着存在一门严格的心的科学。

需要注意的是，假使我们不具有社会和物理现象间的这种相互作用，阻止我们去获得严格社会科学的这个障碍就不存在。请想一想我在前面曾举过的乔姆斯基普遍语法假设的例子。假设在我们个人的头脑中生来就编排着普遍语法规则。这些规则在脑中是与生俱有的，并且又处在有机体与环境的任何关系之外，不管这些规则会怎样相互联系着和怎样的复杂，在我们的论证中从来就没有否认过存在着联系这些规则和脑特征的严格的心理—物理法则。另外，许多动物具有有意识的心理状态，但据我们所知它们不具备与人的语言和社会制度联系在一起的自我涉及。在我的论证中从来就没有否认过一门有关动物行为的科学存在的可能性。例如，可能存在着使鸟的脑状态与它们的筑巢行为相关联的严格法则。

我答应过要至少为你们提供一个逐步深入的论证概要。让我们来看看这个诺言实现了多少，我们把论证按一系列步骤展

开。

1．如果存在着物理学法则意义上的社会科学法则，在用社会和心理词汇确定的现象与用物理词汇确定的现象之间就必须存在着某种系统的关联。这种关联可能像气象现象与物理现象相联系的方式那样复杂，但它必须是某种系统的关联。用现在的行话来说，在较高和较低层次之间必须存在着某些中介原则。

2．社会现象在很大程度上是由人们采取的心理态度规定的。货币、允诺或婚姻之所以是货币允诺或婚姻，在很大程度上是因为人们把它们看作是货币、允诺或婚姻。

3．这就产生了一个推论：这些范畴在物理属性上是无穷无尽的。严格地说，对我们所能认作或约定为货币、允诺或婚姻的东西不存在物理的限制。

4．由此表明，在世界的社会特征和物理特征之间，即在用社会词汇和物理词汇描述的共同现象之间不可能存在任何中介原则。我们甚至不可能具有我们用于气象和消化的那种松散分离的原则。

5．并且，在用心理词汇描述的现象和用神经生理词汇描述的现象之间，即在心与脑之间要得到这类合适的中介原则也是不可能的。这是因为对于任何特定的社会概念，存在着无数种刺激状态。这一极大的范围妨碍了并非生长在我们头脑中的概念以系统关联心理与物理特征的方式被实在化。

在本章结束之际，我想来描述一下我所认为的社会科学的真实特性。一般说来，社会科学涉及的是意向性的种种方面。经济学涉及的是商品和劳务的生产与分配。请看，一个从事经济工作的经济学家完全把意向性视为理所当然的。他假设企业家总设法赚钱，消费者总愿宽裕而不愿拮据。那么这个经济学

"法则"就表明了系统的结果或这样假设的结论。给定某些假设，这位经济学家能推断出精明的企业家会在他们的边际成本和边际收益相等的地方销售。这里要注意，这个法则没有预言销售者会自问："我是在边际成本与边际收益相等的地方销售吗？"也就是说，这个法则不表明个体意向性的内容，它得出的是这种意向性的结果。微观经济学的厂商理论得出的是对用户和行业从事购买、生产和销售的愿望和可能性进行某些假设的结果。宏观经济学得出的是对整个国家和社会进行这样假设的结果。但是一个经济学家不一定要去关心这样一类问题："到底什么是货币？"或者"到底什么是一个愿望？"如果他非常精通福利经济学，他可能会去关心企业家和消费者愿望的准确特性，但即使在这种情况下，从意向性事实中得出的结果仍是他研究领域中的系统部分。

由于经济学不是以像分子结构这样的物理属性的系统事实为依据的，不是像化学那样以分子结构的系统事实为依据，而事实上涉及的是人的意向性、愿望、实践、技术状态与知识状态，因此经济学不能够摆脱历史或环境。经济学作为一门科学以某些关于人与社会的历史事实为假设前提，这些历史事实本身并不是经济学的组成部分。当这些事实发生了变化，经济学也要随之改变。例如，菲利普斯曲线是一个揭示工业社会中一系列因素之间关系的公式，一直到现在，这个公式似乎对工业社会中的经济现实作出了一种精确的描述。但近来它并不那么灵了。多数经济学家认为这是因为它没有精确地描述现实。但他们可以认为，也许这个公式在当时确实曾精确地描述过现实，可是，自从发生了石油危机和七十年代的种种其他事件，现实发生了变化。经济学是一门系统的形式化的科学，但它不能独立于环境或脱离开历史。它建立在人的实践基础上，而这些实践自身不

是一成不变的、永恒的或一劳永逸的。如果出于某种原因，货币不得不使用冰来制造，那么货币在温度高于摄氏零度的时候融化，就是一条严格的经济学法则。但是这条法则也只有在货币必须要由冰来制造的时候才能生效，除此以外，它没有告诉我们任何我们对货币感兴趣的东西。

现在让我们来看看语言学。当代语言学的一般任务是去陈述揭示各种自然语言中声音和意义之间关系的语音的、语法的和语义的种种规则。一种十分完美的语言科学应当对每种自然人类语言都提出一套完整的规则。我并不能肯定这就是语言学的真正目标，我甚至不能肯定这是一个可能达到的目标，但对于我们目前的讨论，重要的是在于强调它也是一门意向性的应用科学。它完全不像化学或地质学。它所关心的是，详细说明操各种语言的人心中那些由历史决定了的意向内容，这些意向内容实际是造成人的语言能力的原因。将语言学与经济学连接在一起的粘接剂是人的意向性。

现在，我们能把本章的结论十分简明地表述出来了。社会与自然科学之间的基本中断不是仅仅出自一种社会与物理现象间断性连接的事实，也不是出自社会学科的基本概念在物理学中没有“共鸣”的事实，甚至也不是由于社会生活的巨大复杂性。许多像地质学、生物学和气象学这样的学科具有这些特征，但这并没有妨碍它们成为系统的自然科学。社会和自然科学之间的根本中断来源于社会和心理现象的内在心理特性。

社会科学靠心推动这一事实是它与自然科学相比之下的弱点所在，但这也恰恰是社会科学的力量所在。我们期望从社会科学中获得的以及我们从最成熟的社会科学中获得的是纯粹的和应用的意向性理论。

六　意志的自由

在以上各章，我力图回答的是一些关于我们作为人类如何与周围宇宙相统一的问题，在我看来这是一些最令人困惑不解的问题。我们关于我们自己作为自由的行为者的概念，对我们全部有关自身的概念来说是基本的概念。从最好的愿望出发，我希望既能够保留我的常识概念又能够保留我的科学信念。例如，对于心身关系问题，我能够做到这一点。但是，一旦涉及到自由与决定论的问题，我与许多其他哲学家一样就不能够使两者统一起来了。

人们总以为这个困扰着人们两千多年的自由意志的难题现在已经得到了最后解决。实际上多数哲学家也认为这个难题已得到了解决。他们认为，它已被托马斯·霍布斯和大卫·休谟以及许多其他经验主义的心灵哲学家们解决了，并且一进入二十世纪他们的解决方法不断得到重复和改进。我认为这个难题依然没有得到解决。在这一讲，我想先为你们描述一下这个难题，然后再回答为什么当代对这个难题的解决不能算作是一种解决，最后我试图说明为什么这个难题会一直存在。

一方面，我们倾向于认为，既然自然界是由微粒以及它们的相互关系构成，并且既然万事万物都可通过这些微粒以及它们的相互关系来说明，那么就根本不存在自由意志的地盘。就人的自由而言，它与物理学是否像牛顿物理学那样是决定论的，还是像当代量子力学那样允许存在一种粒子物理层次上的非决定

性无关。物理学中的粒子层次上的非决定论确实没有给任何自由意志的学说以任何支持。因为首先，粒子层次上的统计的非决定性并不表明在关系到我们的物体层次上，例如人的肉体层次上，也存在着非决定性。其次，即使在物理粒子的行为中存在着一种非决定的成分，也就是说即使它们仅仅在统计上是可预言的，单就这种情况而论仍然没有给人意志的自由留下地盘。因为粒子仅仅在统计上被决定的事实，说明不了人心能迫使这些在统计上决定了的粒子改变它们轨迹的情况。非决定论证明不了存在着或可能存在着某种人类自由的心理之能，这种心理之能能够使粒子朝着不再有其他可能的方向移动。所以，关于物理学我们所知道的一切确实似乎都在迫使我们接受某种人类自由的否定形式。

最形象化地表达这一决定论概念的还要数拉普拉斯的描述：如果一个理想的观察者在给定的一刻知道全部粒子的位置，并且知道全部支配它们运动的法则，他就能够预见和回顾宇宙的全部历史。当代量子力学的拉普拉斯式的一些预言也许是带有统计性的，但这里仍不会给意志的自由留下任何地盘。

关于决定论的主张我们先介绍到这里。现在让我们来看看对于自由意志的论证。正像许多哲学家所指出的那样，如果存在着我们大家都熟悉的经验事实，那么我们自己的选择、决定、推理和筹划似乎对我们实际的行为产生影响这一点就是一个简单的事实。在生活中我们有各种各样的经验，其中我们恰恰有这样一个经验事实：尽管我们已做了某事，但我们完全知道我们本可能去做别的事。我们知道我们本可能去做别的事，因为我们出于种种原因只选择了一件事去做。但是我们意识到如果我们选择做别的事也存在着种种原因，我们确实也可以从这些原因出发去行动，去选择做别的事。换一种方式来表述这种观点就

是：如果把从一个斜面上滚落的物体的行为看成是可预言的话，那么从这个意义上说，我们的行为就是不可预言的，这是一个普通的经验事实。我们的行为之所以在这种意义上是不可预言的，原因在于我们除了事实上所做的以外，我们常常可能去做别的。人的自由仅仅是一个经验事实。如果我们想为这一事实提供某种经验的证明，我们只能进一步达到这样的事实：任何人对我们的行为愿意作的任何预言到我们这里总能被证明是假的。如果某人预言我要去做某事，我恰恰会去做别的事，而对于从山坡上滑下的冰块、从斜面上滚落的球或在自己椭圆轨道中运行的行星则根本不存在这种选择的余地。

这是一个特有的哲学之谜。一方面，一组非常强有力的论证迫使我们接受在宇宙中没有自由意志地位的结论。另一方面，一系列基于我们自己经验事实的强有力的论证又使我们倾向于接受这一结论：必定存在着某种自由意志，因为我们无时无刻不在体验着它。

对这个哲学之谜有一种公认的解决方法。按照这种解决方法，自由意志与决定论是彼此完全相容的。世间的一切当然是被决定的，但是一些人类行动依然是自由的。说一些人类行动是自由的并没有否认它们是被决定的，而只是说它们没有受到强制。我们没有被迫去采取这些行动。而例如，如果一个人在枪口的逼迫下去做某事，或者他正受着某种心理的强迫，那他的行为才真正是不自由的。但是另一方面，如果他自由地行动，如果他就像我们所说的按照他自己的自由意志去行动，那他的行为就是自由的。当然它也是完全被决定的，因为他行为的每个方面都是由作用在构成他肉体的粒子之上的物理力所决定，这正像这些物理力作用在宇宙中所有物体之上的情况一样。所以，自由的行为是存在的，但是它在被决定的世界中仅仅是一个小

小的角落，在这个小小的角落里，人的行为依然是被决定的，只是不存在某些种强制和强迫。

由于这种观点主张的是自由意志与决定论的相容性，人们通常把它简称为“相容论”。我认为这种解决自由意志难题的方法是不恰当的。原因在于，这个自由意志的难题不是关于是否存在导致我们做事的内部心理理由和外部物理原因以及内部强迫的问题。而是这样一个问题：不论我们行为的原因是什么，它们是否足以决定我们的行为使得事情不得不按它们确实发生的方式发生。

这个难题还可用另一种方法表述出来。在所有其他条件保持相同的情况下，说一个人他本可能采取别的行为是永真的吗？例如，假设一个人选定去投保守党的票，在所有其他条件保持相同的情况下他会不会选定去投另外一个党派的票？看来，相容论确实回答不了这个在某种程度上为普通的自由意志概念留有余地的问题。相容论所主张的是，在所有其他条件保持相同的情况下，所有行为都是以本来不可能出现例外的那种方式被决定的。一切发生了的事情当然是被决定了的。只是某些事情是由某些种内部心理原因（我们把它们称为我们的“行动理由”）而不是外部强制和心理强迫决定的。所以，我们仍然面临着一个难题：说一个人他本可能采取别的行为是永真的吗？

那么相容论留下的这个难题也就是它没有回答的这个问题：“在所有其他条件保持相同的情况下，我们有可能采取别的行为吗？”在某种程度上这与我们相信我们自己的自由意志是相一致的。总之，相容论否定了自由意志的实质而只保留了它的语词外壳。

让我们再重新开始我们的探讨。我说过我们确信自己有自由意志，这种确信是建立在人的经验事实的基础之上的。但是

这些经验何以是可靠的？正像我在前面说过的，这种常常由哲学家们描述的使我们倾向于相信我们自己的自由意志的典型情况就是：我们面临着一连串的选择、择其最佳行动的推理、决断以及去做我们已决定去做的事。

我们把这种经验看作是对人的自由学说的一种支持，但也许我们的这种信念是虚幻的。请考虑这样一个例子。一个典型的催眠实验有如下形式：在催眠状态下，病人接受一个催眠后的暗示。例如，你可以让他做一些无关紧要而又无害处的事情，比如叫他在地板上爬来爬去。当这个病人从催眠状态醒来之后，也许是当他在进行交谈、坐在那里休息或喝咖啡的时候，他突然会说出一些类似这样的话来："这屋里的地板多漂亮啊！"、"我想看一下这块地毯"或"我正打算花钱装饰一下地板，我得好好看看这块地板。"说罢，他便开始在地板上爬来爬去。这些情况令人感兴趣的是病人总是为他的行为寻找一些多少还算说得过去的理由。这就是说，在他自己看来，他的行为似乎是自由的。从另一方面来看，我们有充分的理由相信他的行为完全是不自由的。他为他那个在地板上爬来爬去的所谓决定寻找的理由是与他的行为不相干的，他的行为是预先被决定的，事实上他处在催眠后暗示的控制之下。任何一个知道他的这些情况的人都能事先预料到他的行为。于是，提出决定论难题的一种方式或至少是提出决定论难题的一个方面的一种方式就是："人的全部行为都如同这般情景吗？"人的全部行为都如同这个为催眠后暗示所左右的人那样吗？

但是如果我们再认真地分析一下这个例子，它似乎又为自由意志提供了一种论证而不是否证。行为者以为他在自由地行动，而事实上他的行为是被决定的。如果人的全部行为都像这样，从经验上看似乎是极不可能的。有时人们是处在催眠状态

中的，有时人们还受着不可自制的无意识冲动的支配。但是他们总处在这种状态中吗？人的全部行为都是由这种心理的强迫决定的吗？如果我们试图把心理决定论作为一种符合事实的主张，那就大错特错了。心理决定论认为，在先的心理原因就像它们决定着受催眠的病人或吸海洛因成瘾者的行为那样决定着我们的全部行为。在这种观点看来，全部行为总是以这种或那种方式接受着心理的强迫。但是有效的证明表明这种论点是错误的。我们的正常行为确实是建立在我们的意向状态——我们的信念、希望、恐惧和愿望等基础之上的，在这种意义上，我们的心理状态起着原因的作用。但是这种因果形式不是决定论的。我们可以确实具有这些心理状态而不去做我们已做的事。就心理的原因而论，我本可以去做别的。另一方面，催眠状态和在心理上受到强迫的情况通常是属于变态的并且很容易与正常的自由行动区别开来。所以，从心理上说，存在着人类自由的地盘。

但是这种解决方法不正是对相容论的一种推进吗？我们不是恰恰又一次肯定了所有行为是被决定的吗？只不过我们所说的自由行为是那种由理性思想过程决定的自由行为。有时候有意识的、理性的思想过程与那种催眠状态没有什么不同，而有时候它们又与正常状态没有什么不同。在那些正常状态下，我们说行为者是真正自由的。但是，那些正常的理性思想过程当然又像任何其他事物一样是被决定的。所以，我们又一次得出了这样的结论——在我们出生前的千百万年，我们所做的一切都已被完完全全地载入史册，因此从哲学所关切的意义上说，我们所做的一切都是不自由的。如果我们情愿以为我们的行为是自由的，那么这仅仅是一个采用一个传统术语的问题，正像我们明明知道太阳根本不会落下去而继续说“日落”一样，尽管不存在自由意志现象，我们仍可以继续说“按照我们自己的自由意志去

行动。”

为解决这个问题而去考察一种哲学论点或任何其他论点的一个方法就是去问：“如果那种论点为真，世界将会是怎样一种状况？相反如果那种论点为假，世界又会是怎样一种状况？这两种状况会有什么不同？”我相信决定论部分可被人接受之处在于，它似乎与世界的实际进程相一致，至少在我们从物理学中所知的范围内是这样的。这就是说，如果决定论为真，那么世界还是以它所运行的那种方式不断地运行下去，唯一不同的是我们关于世界进程的那种特定的信念就会为假。这些信念对我们极为重要，因为它们涉及到我们的这种信念，即我们本可以以不同于我们实际所做的方式去做事的信念，而这种信念又是与有关道德责任和对我们自己人性的信念联系在一起的。但是如果自由意志论即自由意志的论点为真，那我们就只好对我们关于世界的信念做一些真正根本的改变了。我们为了要获得彻底的自由，似乎就只好去假设我们每一个人都有一个能够干涉作为原因的自然秩序的自我。这就是说，似乎我们不得不有着某种能够使分子偏离它们轨迹的实体。我不知道这样一种观点是否更能用理智来解释，但是这种观点与我们从物理学中所知关于世界如何运行的观点确实是不相容的。我们没有任何证据认为应该放弃物理学的理论而去青睐这样一种观点。

至此，在我们解决决定论与自由意志的信念彼此冲突的努力中，我们似乎变得完全无路可走了。科学不给自由意志以立足之地，物理学中的非决定论不给它提供依据。另一方面，我们又不能放弃自由意志的信念。让我们对这两点再作更深一步的考察。

按照当代科学的观点究竟为什么没有自由意志的位置？在物理学中我们基本的说明途径是采取自下而上的方式。比如说，

我们说明像玻璃的透明性或水的液体性这样一种现象的表面特征活动所依据的是像分子这样的微观粒子活动。心与脑的关系就是这样一种关系的一例。正如我在第一章曾讨论过的，心理特征是由神经生理现象导致的并体现在神经生理现象中。但是，我们从心到身所得出的因果关系是在一段时间内的自上而下的因果关系。我们得出在一段时间内的自上而下的因果关系是因为上层次与下层次是结合在一起的。例如，假设我想把我的运动神经元轴突端板处神经递质乙酰胆碱释放出来，我只要决定反复举起我的手臂就能够做到。在这里，心理事件，即举起我手臂的意向导致了物理事件，即乙酰胆碱的释放。如果说存在着自上而下的因果关系的话，那么这就是一例。但是这种自上而下的因果关系之所以存在是因为心理事件首先是建立在神经生理的基础之上的。所以，与这种对自上而下的因果关系的描述相对应，还有一种对同一系列事件的描述，在这种描述中，因果关系完全是自下向上反弹的，即它们涉及到完全是神经元和突触中的神经元冲动速率等问题。只要我们接受了这个自然界如何运行的概念，那么看来就不会为自由意志留下任何地盘。因为按照这一概念，心只是在作为自然界的一部分的范围内对自然界产生影响的。但如果是这样的话，心就与周围自然界一样，它的特性在其物理学的基本微观层次上就是被决定的。

这是一个在本章中非常基本而又重要的观点，所以我们再重复一遍。最令人困惑的决定论形式不是心理的决定论。这种认为我们的心理状态足以决定我们所做的一切的看法可能恰恰是错误的。令人困惑的决定论形式具有更基本和更重要的形式。由于世界的全部表面特征完全是由微观元素的系统导致并体现在这些系统中的，所以微观元素的活动足以决定所发生的一切事物。这样一种“自下而上”的世界图式容许有“自上而下”

的因果关系(比如，我们的心能影响我们的肉体)。但是自上而下的因果关系之所以能起作用仅仅是因为上层次已被下层次所导致并已体现在下层次之中。

好，让我们再转向下一个显而易见的问题。在我们的经验中是什么原因使我们不可能放弃对自由意志的确信？如果自由是一种幻觉，为什么它是一种我们似乎不能够放弃的幻觉？关于我们人类自由的概念，我们首先要注意的是它与意识在本质上是联系在一起的。我们只是把自由归结为有意识的存在。例如，如果某人制造了一个我们认为是完全没有意识的机器人，我们决不会把它看作是自由的。即使我们发现它的行为是随机的和难以预料的，我们也不会说它是在我们所认为的我们自己自由行动的那种意义上自由地行动。另一方面，如果某人制造了一个使我们确信在与我们同等的意义上具有意识的机器人，那么机器人是否具有自由意志这个问题至少是一个悬而未决的问题。

其次，我们要注意的是，并不是任何意识状态都能为我们提供对人类自由的确信。如果生命完全由被动的知觉感受组成，那么在我看来我们甚至连人类自由的念头都不会出现。如果你把自己想象成是完全不动的，你不仅完全不能够移动甚至还不能决定你自己的思想进程，但你仍能接受刺激，例如你可以有周期性的轻微痛感，那么对于你的这种状态我们没有丝毫理由断定你具有自由意志。

我在前面曾说过，多数哲学家认为对人类自由的确信在本质上以某种方式受到理性抉择过程的制约。但我认为这只是说对了一部分。事实上，对原因的估量，对于给予我们对自由确信的那种经验来说，只是一种极为特殊的情况。使我们确信人类自由的典型经验，即一种使我们无法抹煞确信自由的经验，是

那种进行自愿的、有意向的行动的经验。在我们讨论意向性时，我们集中探讨了行动中有意识意向的意向性形式，按我所描述的那种方式意向性是作为原因的，它的满足条件是一定身体动作的发生，而且身体动作的发生是由行动中那个意向导致的。正是这种经验成为我们确信自由意志的基石。为什么？当你从事正常的和日常的行动时，请你仔细反思一下你所具有的那些经验的特点。你会感到在这些经验中存在着选择行动过程的可能性。当你举起你的手臂、穿过房间或喝水的时候，你会发现在你的经验中随时有一种行动过程的选择感。

如果要用语言来表达观察的经验与行动的经验彼此有什么不同的话，那么观察经验的意思就是："我见此，"而行动经验的意思是："我做此。"但是"我做此"带有"我可能做彼"的意思。在正常的行为中，在我们每做一件事的此时此地，即在所有其他条件保持相同的情况下，我们都带有这种可能去做彼的确信，而不论这种确信是有效的还是无效的。我以为这就是我们对我们自己的自由意志深信不移的根源。强调我现在讨论的是正常的人的行动也许是十分重要的。如果一个人被一种巨大的激情所支配，比如在他大发雷霆时，他就会失去这种自由感，他甚至会对他的所作所为感到震惊。

一旦我们注意到了行为经验的这一特征，我在前面提到的那些大量令人费解的现象就容易得到解释了。例如，接受催眠后暗示的那个人，尽管他可以自认为他在自由地行动，为什么我们认为他不是在我们行动的那种意义上自由地行动？原因在于他在一种至关重要的意义上不知道他在做什么。他那个实际的行动中的意向完全是无意识的。他认为对他很自然的那些选择与他行动的实际动机是无关的。此外，也请注意，相容论例举的"强迫"行为，在许多情况下仍然含有自由的经验。如果某人用

枪逼着我去做事，即使是在这种情况下，我仍具有一种含有行动过程选择感的经验。例如，如果我在枪口的逼迫下走过房间，我仍有一部分这样的体验：在每走一步时我都有可能改变我的行为。对自由的这种体验在任何带有一种意向的行为中都是一个基本的组成部分。

如果我们把行动的正常状态与彭菲尔德状态即刺激运动皮质后所产生的不自觉的上下肢运动加以比较的话，就能确认这种情况。在彭菲尔德状态下，病人被动地体验着运动，就像他们体验着一种声音或一种痛感一样。与有意向的行动不同，在这种经验中不存在选择自由。要清楚地认识到这一点，你不妨去假想你的一部分生活在很大程度上与彭菲尔德实验一样。你不是在走过房间，而只是发现你的身体在移动；你不是在讲话，而只是听到或感觉到从你嘴里说出的语词。如果你把你的经验假想为是一个纯粹被动的然而是有意识的木偶的经验，你就已假想到了排除自由经验的情况。但是在典型的有意向的行动情况下，我们无法排除自由的经验。它是行动经验的一个基本的组成部分。

我相信，这一点也正好说明了为什么我们不能放弃对自由的确信。我们发现一旦我们理解了对太阳系日心说的论证，就能轻而易举地放弃那种大地是一平面的信念。同样，在我们观察日落时，不管出现什么现象都不会迫使我们相信太阳是从地球后面落下去的，我们相信日落现象完全是由地球自转造成的一种假象。对以上每种情况来说，放弃常识性的确信是可能的，因为取代常识的那个假设既首先解释了导致那种确信的经验，又说明了用常识的观点无法解释的许多其他事实。这正是我们放弃了大地是一平面的信念和字面意义上"日落"的信念而去接受哥白尼太阳系概念的原因。但是我们不能照此放弃对自由的

确信，因为这种确信存在于每个正常的、有意识的意向行动中。我们使用这种确信去辨别和说明行动。不管这种自由感是经过预先思考的还是本能自发的都是任何行动的一个组成部分，它不仅仅是深思熟虑的特点，自由感与深思熟虑没有本质的联系，深思熟虑只是一种特例而已。

尽管地球看上去是平的，但我们不能按照这个视地球为平面的假设去航行，然而我们确实是在按照自由的假设去行动的。事实上，不管我们掌握了多少关于世界如何作为一个被决定的物理系统运行的知识，只要没有这种自由的假设，我们就不会有行动。

现在我们可以把包含在这个讨论中的结论提取出来了。首先，如果对决定论的怀疑是由于担忧我们的全部行为事实上是受到心理强迫的，那么这种担忧是没有根据的。只要心理的决定论与其他假设一样是一种经验的假设，那么从我们目前所能得到的证据来看它就是错误的。因此，这为我们提供了一种经过修正的相容论的形式。它告诉我们心理的自由意志论与物理的决定论是彼此相容的。

其次，它甚至给予我们一种“可能”感，人们的行为尽管是被决定的，但正是在那种“可能”的意义上他们有可能采取别的行动。人们有可能采取别的行动，其意义仅仅是就心理因素而论的。关于什么是我们能够去做的和什么是我们有可能去做的这个能力的概念常常涉及到某些这样的标准：例如，尽管我在1980年的美国大选中没有投卡特的票，我有可能投他的票，但是我决没有可能去投乔治·华盛顿的票，因为他不是一个候选人。所以，在一种“可能”感中具有一个可供我去选择的范围，并且在这种意义上，在所有其他条件相等的情况下，有许多我本可以去做而没有做的事。与此相似，由于作用于我的心理因素一般不会

总强迫我以奇特的方式去行动，所以从心理上说，我常常有可能去做不同于我事实上所做的事。

第三，但是这种形式的相容论仍没有为我们提供我们要求用彻底的自由意志论来解决自由与决定论之间冲突的任何东西。三百年来的科学是建立在“自下而上”这个物理说明的概念之上的，只要我们接受了这一概念，那么关于我们自身的心理事实也同任何其他较高层次的事实一样，其原因完全可以用基本的微观物理层次上的元素系统来解释，并且这些心理事实是在基本的微观物理层次上的元素系统中实现的。我们关于物理实在的概念根本不考虑绝对的自由。

第四，也是最后一点，由于某些我还没有真正理解的原因，进化赋予了我们一种随意行动的经验形式，在这里，自由的经验，即那种带有选择可能感的经验，是存在于那种有意识的、随意的和有意向的人类行为的结构之中的。由于这个原因，我相信无论是我们的这个讨论还是任何其他讨论都将永远不会使我们确信我们的行为是不自由的。

在本书中，我的目的是去描绘这样两个概念之间的关系：一个是我们把我们自己作为理性的、自由的、有意识的和自觉的行为者的概念，一个是我们把世界视作由无意识、无意义的物理粒子构成的概念。我们很容易认为，正像我们已发现大部分常识不能恰当地描绘世界真正在如何运行一样，我们也可能发现关于我们自己和我们行为的概念完全为假。但是这种可能性是有限的。对实在与现象的区分不能套用在意识的存在上。因为，如果在我看来我是有意识的，那么我**就是**有意识的。我们可能发现我们自己和我们行为的种种令人惊叹的东西，但我们不可能发现我们不具有心灵以及心灵不包含意识的、主观的和意向性的心理状态；我们也不可能发现我们不再试图去完成的自发

的、自由的和意向的行动。我自己设定的这个难题不是去证明这些东西的存在，而是去考察它们的状况以及它们对于我们自然界其他部分的概念的含义。除了某些重要的题外问题，我讨论的主题始终是：我们关于我们自己的常识性的心理主义概念与我们把自然界视作一个物理系统的概念是完全一致的。

附　　录

什么是言语行为?[1]

一、引　言

在一种包含说者、听者和说者所作的表述在内的典型言语境况中，有许多行为与说者的表述联系在一起。从特征上看，说话者将移动他的颌、舌，并且要发出声音。此外，从特征上看，他将完成像通知、激怒和打搅他的听者这样一类行为；从特征上看，他将进一步完成像指称肯尼迪、赫鲁晓夫或北极这样一类行为；而且他也将完成像作陈述、提问题、下命令、作报道、问候和警告这样一类行为。这最后一类中的行为就是奥斯汀所说的非表现行为[2]。在本文中，我要讨论的就是这类行为，所以又可以把本文题目叫作“什么是非表现行为?”。我不打算去定义“非表现行为”这个表达式，尽管我对一种特殊的非表现行为的分析如果是成功的话就可能为一种定义提供依据。与非表现行为相联系的一些英语动词和动词短语有：陈述、断定、描述、警告、议论、评论、命令、指令、请求、批驳、道歉、指责、批准、欢迎、允许、表达赞成和表达遗憾。奥斯汀认为，在英语中存在着一千多个这样的表达式。[3]

① 选自马克斯·布莱克编，《美国哲学》(艾伦和昂温出版社，1965年)，第221—239页。

② 奥斯汀，《如何用言语行事》(牛津，1962年)。

③ 非表现行为(illocutionary act)又译“以言行事行为”，该词由奥斯汀首先使用，指说话者使用语词完成某种行为并带有一种以言行事的力量，它区别于表现行为(locutionary act)和超表现行为(perlocutionary act)。——译者

言语行为有时也被称为语言行为，通过这个引言，我也许能够说明为什么我认为研究言语行为在语言哲学中具有重大的意义。我认为在任何语言交流的模式中都必须包含有一个语言行为。语言交流的单位不是通常人们认为的符号、语词或语句，甚至也不是符号、语词或语句的标记，构成语言交流的基本单位是在完成言语行为中给出标记。更确切地说，在一定条件下给出语句标记就是非表现行为，非表现行为是语言交流的最小单位。

我不知道如何去证明语言交流必然包含着行为，但我能想到一些论据，人们试图通过这些论据来使对此持怀疑态度的人接受这个观点。其中一个论据是叫持怀疑态度的人注意到这样的事实：当他发出一个声音或在纸上写出一个记号使之成为一种语言交流的事例或者说是作为一种消息时，包含在他如此发声和写作中的一个事实就是，他应当把他所发出的声音和写出的记号视为是由一个具有特定意向的人给出的。他不能只将它视为像一块石头、一挂瀑布或一株树那样的自然现象。为了把它视为一种语言交流的事例，就必须假设它的给出就是我所说的言语行为。例如，当代试图破译玛雅人象形文字的一个逻辑的先决条件就是，我们至少假设了我们在石头上见到的这些记号是由多少与我们有些相像的人类创造的，并且是带着某些意向创造出来的。如果我们确信这些记号比如说是由水的侵蚀造成的，那么关于破译这些记号的问题，甚至称它们为象形文字的问题就不会产生了。用语言交流的范畴来解释它们，就必然要把它们的写作作为言语行为来解释。

完成非表现行为就是采取一种由规则支配的行为方式。我将指出像提问、陈述这样的受规则支配的行为方式非常类似于像棒球中的击球、象棋中的跳马那样的受规则支配的行为方式。因此，我打算通过陈述一套完成一个特殊种类的非表现行为的

必要和充分条件，来解释非表现行为的概念，并且为使用标记那种非表现行为表述的表达式（或句法手段）引出一套语义规则。如果我对一种非表现行为的条件和相应规则的陈述是成功的，那么这就会为我们分析其他种行为进而解释一般意义上的非表现行为概念提供一种范型。但是，为了要去实际陈述完成一个非表现行为的条件并引出其规则，我就不得不去讨论其他三个初始概念：规则、命题和意义。我将把对这些概念的讨论限制在为实现本文的主要目的而不可缺少的那些方面上。但即使如此，如果让我把每一个概念需要加以说明的方面统统说完的话，那么对每一个概念的说明都需要写作一篇论文。不过，出于对论述范围的考虑，有时这种详尽的说明是应该放弃的，因此我的说明将是简明扼要的。

二、规　则

近年来，在语言哲学中围绕表达式使用规则的概念所展开的讨论值得重视。一些哲学家甚至说，知道语词的意义仅仅是一个知道它的使用规则的问题。这些讨论有一个令人焦虑的方面：至少据我所知，从没有一个哲学家对哪怕是一种表达式的使用规则给出任何一种恰当的说明。如果意义是一个使用规则的问题，那我们就确实应该能够从解释表达式意义这方面去陈述使用表达式的规则。另一些哲学家也许被他们同行们构造规则的失败弄得灰心丧气，他们对将意义视作规则问题这种时髦的观点予以否定，并且断言根本不存在那种句法规则。我倾向于认为，这种怀疑主义的断言下得过早，并且提出这种断言是由于没有成功地区分不同种类的规则，对此我现在将试图加以说明。

我区分了两类规则：一些规则用于调节先前存在的行为形式。例如，礼仪规则用于调节人与人之间的关系，而这些关系是独立于礼仪规则而存在的。还有一些规则不仅调节而且创造或规定新的行为形式。例如，橄榄球的规则不仅仅调节了橄榄球比赛而且还创造了这种运动的可能性并对这种运动加以规定。橄榄球运动是由符合于这些规则的行为构成的，离开这些规则，橄榄球运动就不会存在。我把这后一类规则称为构成规则，把前一类规则称为调节规则。调节规则调节一种预先存在的活动，这种活动的存在在逻辑上独立于规则的存在。构成规则构成(同时也调节)一种活动，这种活动在逻辑上是依赖于规则的存在而存在的①。

调节规则的特有形式是祈使句，或者说能被释义为祈使句形式，例如："切食物时要用右手拿刀"、"官员们进餐时要系领带"。一些构成规则则采取与之完全不同的形式，例如，在象棋比赛中，如果王在对方的攻击下无路可走就是被将死了；在橄榄球比赛中，当一名带球队员越过对方底线，就是一次底线得分。如果我们的规则范式仅限于祈使句的调节规则，那么那种非祈使的构成规则就会使我们难以理解，甚至会把它们排除在规则之外。需要注意的是，构成规则在特征上几乎是重言式的，因为这种"规则"所提出的东西是"将死"或"底线得分"的不完全定义。当然，这种准重言式的特性是它们作为构成规则的一个必然结果：底线得分的规则必须像按橄榄球的规则定义"橄榄球"那样去定义"底线得分"的概念。例如，一次底线得分可按这样或那样的方式来得分，并且记6分有时是作为一种规则出现的，有时是

---

① 关于这种区分见J·罗尔斯，"两种规则概念"，载《哲学评论》，1955年；约翰·塞尔，"如何从'是'得出'应该'"，载《哲学评论》，1964年。

作为一种分析真理出现的，而且把它看作是一种重言式也就暗示了该规则是一种构成规则这一事实。调节规则一般具有"按X做"或"如果Y则按X做"的形式。在构成规则的集合中，有些规则具有这种形式，但有些也具有"将X定为Y"的形式①。

是否看到这一点，在哲学上事关重大。例如，有些哲学家问："一种允诺怎么能建立一种契约?"这个问题好像在问："一次底线得分怎么能得6分?"这两个问题只能用陈述一个"将X定为Y"形式的规则来回答。

我倾向于认为一些哲学家在陈述表达式使用规则上的失败与另一些哲学家之所以对任何这类规则的存在抱有怀疑主义态度，部分原因至少是由于他们没有认识到构成规则与调节规则之间的区别。在多数哲学家心目中，规则的范式或模型是那种调节规则式的。如果有人要在语义学中寻找纯粹的调节规则，那么他决不会找到任何对逻辑分析有价值的东西。毫无疑问，存在着像"人不应在正式场合下说猥亵话"这种形式的社会规则，但这种规则在解释关于一种语言的语义学中几乎不起什么重要作用。接下来的一个假设是：一种语言的语义学被视为一系列构成规则的系统，并且非表现行为就是按照这种构成规则完成的行为。本文的目的之一就是将某一种言语行为的一套构成规则公式化。如果我所说的这些关于构成规则的内容是正确的话，那么我们也毋须为不是任何规则都采取祈使规则的形式而大惊小怪。我们将确实看到这些规则属于几个不同的范畴，而且没有一个与礼仪规则相像。陈述非表现行为规则的努力也可以被视为是对在言语行为的背后存在着构成规则这个假设的一种检验。如果我们不能给出任何令人满意的规则公式，那么我

① 公式"将X定为Y"最初是由马克斯·布莱克向我提出来的。

们的失败就可以被看作是为推翻这个假设而提出的部分有待于进一步证实的证据。

三、命　　题

在不同的非表现行为之间常常具有共同的特征。请考虑以下语句的表述：

(1) 约翰将离开这个房间吗？

(2) 约翰将离开这个房间。

(3) 约翰，离开这个房间！

(4) 但愿约翰离开这个房间。

(5) 如果约翰打算离开这个房间，我也打算离开。

这些在一特定场合中的每一种表述，从各自的特征来看完成的是不同的非表现行为。从其特征上看，第一句是一种疑问；第二句是关于将来的一种断言，即是一种预言；第三句是一种请求或命令；第四句是对一种愿望的表达；第五句是对意向的一种假设的表达。尽管如此，从特征上看，在完成以上每一种行为时，说话者总会完成这五种非表现行为共同具有的某些附属行为。在每句话中，说话者都指称到约翰这个特定的人，并且都断言到这个人离开房间的行为。这决不能代表他的全部行为，但在每一句表述中这都是他所做的一个组成部分。因此，我将指出，在每一种表述中，尽管非表现行为不同，但至少某些表现行为的指称与断言是相同的。

在我看来，在以上每一种非表现行为中，对某个人(约翰)的指称以及对有关他同一件事的断言就是每一句表述的共同内容。某种可用"约翰将离开这房间"这个从句表达的东西看来就是它们的一个共同特征。我们采用一种把这个共同特征分离

出来的方式，就能毫不歪曲地写出每个语句："我断言约翰将离开这个房间"、"我问是否约翰将离开这个房间"等等。

由于找不到更合适的词，我建议把这种共同的内容称为命题。在从(1)到(5)每一语句的表述过程中，说话者都表达了约翰将离开房间这个命题，我将通过对这种情况的分析来描述这些非表现行为的特征。请注意，我没有说是语句表达了命题，我不知道语句如何能完成那种行为。但我要说的是在对语句的表述中说话者表达了命题。还需注意的是，我把一种命题与对那个命题的一种断言或陈述区别开来。这五个表述都表达了约翰将离开房间这个命题，但只有第二句话才对那个命题进行了断言。一种断言是一种非表现行为，但一种命题则根本不是一种行为，尽管表达一种命题的行为是完成特定的非表现行为的一部分。

对此可以概括地说，我区分了非表现行为与非表现行为的命题内容。当然，并不是每个非表现行为都具有一个命题内容，例如，一声"乌拉！"或"哎哟！"就没有命题内容。在这种或那种著述中早已有过这种区分，像弗雷格、谢弗、刘易斯、赖欣巴赫和黑尔等等许多人曾以不同的方式对此有过论述。

从语义的观点来看，我们能把语句中命题的表示者与非表现力的表示者区分开来。这就是说，对大多数被用于完成非表现行为的语句来说，从分析的目的出发，我们能够说这种语句具有两个(无需分离的)组成部分：表示命题的部分与表示功能的手段。① 表示功能的手段表明怎样去使用命题，或者换句话说，它表明什么是这个表述所具有的非表现力，即说话者在作语句

① 在"我答应我要来"这个语句中，功能表示手段与命题部分是分离的。"我答应来"这个语句与上面的语句意思相同，并且是通过特定的转换从第一句派生而来的，在这个语句中，两个部分不是分离的。

表述时完成的是什么样的非表现行为。在英语中，表示功能的手段包括词序、重音、语调升降曲线、标点、动词语气以及一套所谓完成行为式动词，即我可以在语句的开头使用“我道歉”、“我警告”、“我指出”等等来表示我要完成的那种非表现行为。在实际的言语境况中，语境常常使表述中的非表现力变得很清楚，而不必求助于适当的表示功能的手段。

如果这种语义上的区分是真正重要的区分，那么很可能在句法上也会有某种类似的区分。近年来在转换语法方面的某些进展倾向于支持这种观点。在语句下面的短语标记中，那些相当于表示功能手段的部分与那些相当于命题内容的部分之间存在着一种区别。

这种对表示功能的手段与表示命题的手段的区分将会对我们分析一种非表现行为大有帮助。由于所有各种非表现行为可以共同具有同一命题，所以我们能够把对命题的分析从我们对各种非表现行为的分析中抽出来，放到一边。我认为存在着表达命题的规则，存在着像指称和断言这类东西的规则，但这些规则能够与表示功能的规则分开单独被讨论。在本文中，我将不讨论命题规则而集中讨论使用某些表示功能的手段的规则。

四、意　　义

从特征上看，言语行为是在发出声音或作出标记的过程中完成的。仅仅发声或作标记与完成一种言语行为有什么区别呢？一个区别是，在完成一种言语行为中人发出的声音或作出的标记在其特征上被看成是具有意义的；与此相关的第二个区别是，从其特征上看，人通过这些声音或标记意谓某种东西。从其特征上看，人说话就是通过他所说的东西来意谓某种东西，而

且他所说的东西，即他发出的一串词素，其特征就是具有意义。附带提一句，这一点正是我们不可将完成言语行为与做游戏相类比之处。象棋比赛中的棋子在特征上不具有意义，而且当人移动棋子时在特征上也没有通过这种移动意谓什么。

但是，说话者通过他所说的话去意谓某种东西这是什么意思？某种东西具有意义是什么意思？要回答第一个问题，我建议去借用并修正保罗·格赖斯的某些概念。在标题为"意义"的一篇文章中[①]，格赖斯对"意义"概念的一种涵义作了如下分析：说A通过X意谓某种东西就是说"A试图通过听众对他意向的领会使说出的X在听众中产生某种效果"。在我看来，这对意义的分析是一个有益的开端。首先，因为它表明了意义概念与意向概念之间的密切关系。其次，因为他抓住了我以为是说一门语言的根本性的东西：在说一门语言时，我通过让听者领会我所要传达这些事情的意向来试图把事情传达给听者。例如，从特征上看，当我作出一断言时，我是在力图传达并使我的听者相信某一命题的真理，而且我所使用的方法就是发出某种声音，这种声音就是我企图在听者身上产生预想效果的话语，而这种效果是通过让听者领会我要产生那种效果的意向实现的。我举一例加以说明。我试图让你相信我是法国人，一方面我可以通过每时每刻讲法语、按法国风格穿着打扮、表现出对戴高乐的极端崇拜、与法国人相识交往等等来让你相信我是法国人；另一方面我还可以仅仅通过告诉你我是法国人来让你相信我是法国人。那么，这两种试图让你相信我是法国人的方式有什么区别呢？一个关键的区别就是，在第二种情况中，我通过让你领会我要让你相信我是法国人这个有目的的意向来试图使你相信我是法国

① 《哲学评论》，1957年。

人。这正是告诉你我是法国人所包含的东西之一。当然，如果我用我所描述过的那种行为来让你相信我是法国人，那么这就不是那种通过让你领会我的意向来使你相信我是法国人的方法。我想，在这种情况下即使你领会了我的意向，但你总会存有相当大的疑虑。

可是，不论这种对意义的分析多么有价值，在我看来，它仍在某些方面存在着缺陷。首先，它没有把一个人打算在他的听者身上产生的不同种类的结果，即超表现与非表现区别开来，也没有表明这些不同种类的结果与意义概念相联系的方式。第二个缺陷是，它没有理解意义在什么程度上是一个规则或惯例的问题。也就是说，他对意义的这种解释没有表明一个人通过他的话语意谓某事物与他的话语在语言中实际具有的意思这两方面之间的联系。为了搞清这一点，针对意义的这种分析我提出一个反例。反例中的观点将阐明说话者所意谓的东西与他说出的词所意谓的东西之间的联系。

假设我在第二次世界大战期间是一名美国士兵，并且我被意大利军队俘虏了。假设我希望让意大利军队相信我是一名德国军官，好让他们释放我，我希望能用德语或意大利语告诉他们我是一名德国军官。但让我们假设我的德语或意大利语水平还达不到这种程度，所以，按照我的程度，我企图通过背诵我还记得的一点儿德语来假装作一种告诉他们我是一名德国军官的表述，并且我深信他们的德语水平不足以识破我的计划。假设我只知道一句德语，这还是来自我高中德语课上学的一首诗。因此，我这个被俘的美国人对意大利俘获者说了如下语句："你可知道那柠檬花开的地方？"现在，让我们用格赖斯的术语来描述这种情景。我打算在他们身上产生某种效果，即相信我是一名德国军官的效果，而且我是通过他们对我意向的领会来打算

产生这种效果的。我打算让他们认为我在试图告诉他们我是一名德国军官。但是从这种描述中，也就是当我用德语说"你可知道那地方……"等等的时候就能得出我所意谓的"我是一名德国军官"吗？不仅得不出，而且在这种情况下，用我说出的这句德语去意谓"我是一名德国军官"甚至去意谓"我是一名德国军官"这句德语是完全虚假的，因为这句话的意思是："你可知道那柠檬花开的地方？"当然，我想让我的俘获者误以为我说的意思是："我是一名德国军官，"但构成这一欺骗的一部分就是让他们去思考我说出这句话的德文意思。维特根斯坦在《哲学研究》中有一处说："说'这里冷'，又意谓'这里热'。"① 我们之所以不能做到这一点是因为我们所能意谓的东西是我们正在说的东西的一个功能。意义不仅是一个意向的问题，它也是一个惯例的问题。

我们能够将格赖斯的说明加以修正然后去处理这类反例。在这里，我们已经有了这样一种情况：我通过让别人去领会我要产生某种效果的意向来试图产生某种效果。但是我用来产生这种效果的手段，凭借支配使用这种手段的规则，通过惯例才能被用来作为一种产生不同的非表现效果的方法。因此我们必须按照这样一种方式来修改格赖斯对意义的说明以便使它更清楚：当一个人说什么时，他所意谓的东西决不仅仅偶然地与在他使用的语言中那个语句所意谓的东西相联系。在我们对非表现行为进行分析时，必须抓住意向与惯例这两个方面，同时要特别注意它们之间的关系。在完成一种非表现行为时，说话者通过让听者领会他打算产生某一效果的意向来试图产生某一效果，而且只要他在用词说话，他就会通过这一事实来打算让别人领会

① 《哲学研究》(牛津，1953年)，第510段。

他的意向：他所说出的这些表达式的使用规则将表达式与它效果的产生联系在一起。在对非表现行为的分析中，我们需要表达的正是要素间的这种结合。

五、怎样作允诺

现在，我将对作允诺这种非表现行为进行分析。要进行这种分析，我就要问在一给定语句表述中完成允诺行为的充分必要条件是什么。我将把这些条件当作一套命题来陈述，通过对这些命题的陈述来试图回答这个问题，这些命题的结合蕴含一个说话者作允诺的命题，同时说话者作允诺的这个命题又蕴含这种结合。因此，每一个条件对于完成允诺行为都是一个必要条件，并且将这一套条件放在一起就将是完成这一行为的一个充分条件。

如果我们得到这样一套条件，我们就能从中引出一套功能表示手段的使用规则。这种方法类似于在一个人知道正确地跳马、挪车、将军等等情况下通过问自己什么是下棋的充分必要条件去发现下棋的规则。这就好比我们知道如何下棋，却不知道如何表述下棋的规则并且希望把这些规则表述出来。我们知道如何去做非表现行为的游戏，但一般来说它们是在没有明确的规则公式情况下完成的。要获得这种公式的第一步就要提出完成一种特殊非表现行为的条件。因此，我们的探究将为双重的哲学目的服务。通过陈述一套完成一种特殊非表现行为的条件，我们将对非表现行为概念作出部分解释，并且这也将为下一步提出其规则的公式铺平道路。

我感到陈述这些条件是一件很困难的事，我对我将要列出的这几条是很不满意的。其所以困难的一个原因是允诺概念像

日常语言中大多数概念一样没有十分严格的规则。有各种各样的奇特、异常和模棱两可的允诺，而且人们能举出一些多少有些稀奇古怪的反例来反驳我的分析。我倾向于认为，我们将不能够获得一套完全反映“允诺”一词日常用法的敲定的充分必要条件。因此，我将把我的讨论限制在允诺概念的中心，省略次要的、模棱两可的和有部分缺陷的情况。我也把我的讨论限制在充分明确的允诺上，省略通过短语、暗示、隐喻等省略手段作出的允诺。

另一个困难来自我不想以某种循环的形式来陈述这些条件。我想要给出的完成某一非表现行为的一系列条件，它们不应提到任何非表现行为的完成。我要想对一种非表现行为的概念提出一种一般意义上的解释就需要满足这个条件，否则我就应该去说明不同非表现行为之间的关系。可是，尽管在分析中将不指称非表现的行为，但某些非表现的概念将出现在分析对象和分析者中。我认为，由于构成规则的性质，这种循环形式是不可避免的。

在描述这些条件的过程中，我将首先考虑真诚允诺的情况，然后说明这些条件经过如何修改就能适用于虚假的允诺。由于我们的探究是属于语义学的而不是语法学的，我将仅仅假设存在语法上符合规范的语句。

设定听者H在场时，说话者S说出语句T，而且，在说出T时，S真诚地（而且无缺陷地）对H作出允诺P当且仅当：

(1) 获得正常的输入与输出条件。我使用“输入”与“输出”两词来划定这种大而无限定的条件范围，在这种条件下任何认真的语言交流都是可能的。“输出”包含可使人听懂的说话条件，“输入”包含理解条件。把它们合在一起就包含了这样的条件：说者与听者都知道怎样说这种语言；双方都意识到他们在做

什么；说话者不是在强迫和恐吓之下说话；他们没有像耳聋、失语症、喉炎这样的妨碍交流的生理障碍；他们的话语也并不是在开玩笑或说笑话。

(2) 在说出T时，S表达P。这个条件把命题内容与言语行为的其他部分分开，这样能使我们在对其他部分的分析中集中分析允诺的特性。

(3) 在表达P时，S断定S的将来行为A。在作允诺时，表示功能的手段是一种范围包括某些命题特征的表达式。在允诺中，说话者必须对他的行为进行断言，并且他所断言的行为不能是他过去的行为。我不能对已经完成的事情作出允诺，也不能对别人将要做的事情作出允诺（尽管我能对我看见他将要做什么作出允诺）。出于当前的目的，我把一种行为的概念看成是包括有抑制行为、完成的系列行为，还可以包括状态和状况：我可以允诺不做某事，我可以允诺反复地做某事，我还可以允诺进入或保持某种状态或状况。我把条件(2)和(3)称为命题内容条件。

(4) H愿意S做A而不愿意他不做A，并且S相信H愿意他做A而不愿意他不做A。允诺与恐吓两者之间的一个至关重要的区别就是，允诺是保证为你而不是对你做某事，而恐吓是保证对你而不是为你做某事。一种允诺，如果它所允诺的东西是接受允诺的人不希望去做的事情，那么这种允诺就是有缺陷的；同样，如果允诺者不相信接受允诺的人希望让那件事做成，那么这种允诺也是有缺陷的；因为一种没有缺陷的允诺必须有心去做允诺而不是去恐吓或警告。为避免明显的反例，我认为这个分成两部分的双重条件是必要的。

可是，人们可以想出一些在表面上与我所陈述的这个条件相悖的例子。假设我对一个懒惰的学生说："如果你不按时交作

业，我保证让你这门课不及格。"这是一种允诺吗？我想不是的，把它视为一种警告倒更自然一些，也许还可以把它视为一种威胁。但是在这种情况下为什么可以使用"我保证"这种表达方式呢？我认为我们之所以在这里可以使用它，是因为"我保证"和"我特此保证"在英语中属于作保证的最强的表示功能的手段。由于这个原因，我们常常在完成有些言语行为时使用这样的表达式，这些言语行为并不是在严格地作允诺，而只是为了强调我们所作的保证。为了弄清楚这一点，让我们从不同途径来考虑另一个对我们的分析提出的所谓反例。有时当人们作一种强调性的断言时，可以听到人们说"我保证"，我觉得这种情况在美国比在英国更普遍些。例如，假设我指责你偷了钱。我对你说："这钱难道不是你偷的吗？"你回答说："不是我偷的，我向你保证，我没有偷。"在这种情况下，你作了一种允诺吗？我觉得把你说的这句话视为一种允诺是非常牵强附会的，把这句话看作是一种否定的强调倒更为合适。我们可以把"我保证"这种功能表示手段解释为从真正的允诺中派生出来的东西，在这里它是用于加强你的否定的一种表达式。

条件(4)所陈述的观点大体是说，如果一种有意义的允诺是无缺陷的，那么它所允诺的东西必须是听者要想去做的事情，或者是听者认为符合他的利益的事情，或者是他愿意让别人去做的事情而不是相反的事情，等等；说话者必须对此有所意识、相信、知道，等等。我认为要得到这个条件更完美、更准确的公式需要引入技术上的术语。

(5) S与H显然都不清楚，S在事件的正常进程中将去做A。许多不同种类的非表现行为有一个共同的条件，其大意是说非表现行为必须要有一种要旨，而条件(5)就是这个共同条件的一个实例。例如，如果我请求某人去做他显然已经正在做的或

正要去做的事情，那么我的请求就是无要旨的，而且从这种意义上说我的请求也是有缺陷的。在实际的言语境况中，听众由于懂得完成非表现行为的规则，他们会假定这个条件是被满足的。例如，假设在一次公开演说中，我对我的一个听众说："史密斯，请看这里，注意我在说什么。"为搞清这句话的意思，听众将会假定史密斯一直没有注意听演说或者不管怎么说他是否注意这一点是不明确的，这也就是说这里以某种方式提出了一个他是否注意的问题。这是因为请求的一个条件就是，听者现在正在做还是正要去做被请求做的事情这一点并不是显而易见的。

允诺的情况与此相似。我允诺去做我显然要去做的事，这是不妥的。如果我确实要作这样的允诺，那么我的听众能理解我的话语的唯一方式就是，假定我相信我将要去做这件允诺过的事情这一点是不明显的。一个有着美满婚姻的丈夫对妻子作允诺说下周他将不会抛弃她，这很可能会给她带来更多的忧虑而不是安慰。

此外，我认为这个条件是齐普夫定律陈述的那种现象的一个实例。我认为在我们的语言中，正像人的大多数行为形式一样，有一种最小努力的原则在起作用，在这种情况下，最大的非表现结果的原则以最小的语音行为为结束。我以为条件(5)是它的一个实例。

我把条件(4)与(5)称为准备条件。它们是恰当允诺的先决条件，但它们还没有表明允诺的基本特征。

(6) S 打算做 A。真诚与不真诚允诺之间的最重要的区别在于，在真诚的允诺中说话者打算将允诺付之行动，在不真诚的允诺中说话者不打算将它付之行动。在真诚的允诺中说话者也相信将允诺付之行动（或避免这一行动）是可能的。我认为

他打算将允诺付之行动的这个命题已经蕴含着他认为将允诺付之行动(或避免这一行动)是可能的这个命题。所以，我不把后一个命题作为一个额外的条件来陈述。我把此条件称为真诚条件。

(7) S打算通过说出T使他承担做A的义务。允诺的基本特征是去承担完成某一行为的义务。我认为正是这一条件将允诺(以及同一族的其他成员，如起誓)与其他种类的言语行为区别开来。我们注意到，在对这个条件的陈述中我们只规定了说话者的意向，在以下条件中我们将说明这种意向是如何实现的。不管怎么说，具有这种意向是作允诺的一个必要条件，这一点是清楚的。因为，说话者如果能证明在一给定表述中他不具有这种意向，那么他也就证明了这种表述不是一种允诺。例如，我们知道匹克威克先生没有允诺与那个女人结婚是因为我们知道他没有这方面的意向。

我把此条件称为基本条件。

(8) S打算通过说出T使H产生这样的信念，条件(6)与(7)就是靠对产生这种信念的意向的领会获得的，而且S打算通过H对在惯例上用来产生这种信念的语句的领会来使他的那种意向得到领会。这个条件把我们关于修正格赖斯观点的分析用到了说话者作允诺这个问题上。说话者打算产生某一非表现效果。他是通过让听者领会他打算产生这种效果的意向来达到的。而且，由于他所说出的词项，其词汇和语法特性通过惯例与其产生的效果相联系，所以他也打算通过这一事实来使他的意向得到领会。

严格地说，这个条件可以作为条件(1)的一部分来加以阐述，但由于它在哲学上具有重要意义，所以值得对它单独进行表述。由于以下原因，我感到这样做，有些困难：如果我最初向格

赖斯提出的反对意见确实是有效的，那么一个人完全可以说所有这些重复的意向都是多余的，所需要的只是说话者应该认真地说出一个语句。所有这些效果的产生不过是一个听者所掌握的有关那个语句含义的知识问题，说到底这不过是一个他的语言知识的问题，而听者所具备的这种语言知识正是说话者一开始就假定的。我认为对这种异议的正确答复是条件(8)，它解释了说话者“认真”说出语句，即不仅说出而且还有所意谓，是什么意思。但无论是对这种异议还是对这种异议的答复，它们各自到底有多大的说服力，我是没有完全把握的。

(9) S 和H 所讲言语的语义规则是，T 被正确而真诚地说出，当且仅当获得条件(1)—(8)的语义规则。此条件旨在说明被说出的语句是通过语言的语义规则用来作允诺的语句。将此条件与条件(8)合在一起就能排除上面所举的像被俘士兵这类反例。究竟这些规则的公式是怎样的，我们马上就会看到。

至此我们考虑的只是真诚允诺的情况，但不真诚的允诺依然是允诺，我们现在需要把这种情况考虑在内来看看如何修改允诺的条件。在作不真诚允诺时，说话者不具有作真诚允诺时的全部意向和信念。可是，他声称具有这些意向和信念。的确，正是因为他声称具有他其实没有的意向和信念，我们才把他的行为视为不真诚的。所以，考虑到不真诚的允诺，我们需要指出说话者要对他的信念与意向负责而不像原来只陈述说话者实际具有这种意向和信念这样来修改我们的条件。有这样一个事实可以看出说话者要对他所作的允诺负责，即他不可能真实地说出像这样的话：“我答应做A，但我不打算做A。”说“我答应做A”就是要对打算做A负责，这才抓住了这句话语是否真诚的条件。考虑到不真诚允诺的可能性，我们必须对条件(6)加以修正，使这个条件不光陈述说话者打算做 A，而且还要表明他要对打算

做A负责。为避免循环出现，我用如下言语来表达：

条件(6)补充：S打算通过说出T来使他对打算做A负责。经过这样修改[而且由于我们从分析对象和条件(9)中放弃了"真诚地"一词]，在关于允诺是真诚还是不真诚的问题上，我们的分析就成为中性的了。

六、功能表示手段的用法规则

接下来，我们的任务是从我们这一套条件中引出一套功能表示手段的用法规则。很显然，并不是所有这些条件对于完成这一任务都具有同等的意义。(1)和(8)、(9)形式的条件一般用于所有各种正常的非表现行为，而不是允诺所特有的。用于允诺的功能表示手段规则将相应地建立在条件(2)—(7)之上。

允诺的功能表示手段P，其用法的语义规则如下：

规则1 只有在语句的语境中(或谈话的进一步展开中)才能说出P，该话语断定说话者S的某个将来的行为A。我把此规则称为命题内容规则，它从命题内容条件(2)和(3)中得出。

规则2 只有当听者H愿意S做A而不愿意他不做A，并且S相信H愿意他做A而不愿意他不做A时，才能说出P。

规则3 只有当S在正常的事件进程中做A这一点不是对S和H都很清楚时，才能说出P。我称规则(2)和(3)为准备规则。它们从准备条件(4)和(5)中得出。

规则4 只有当S打算做A时，才能说出P。我把此规则称为真诚规则。它从真诚条件(6)中得出。

规则5 把说出P当作承担做A的义务。我称此规则为基本规则。

这些规则是有序的：只有当规则1被满足时，规则2—5才适用；并且只有当规则2和3也被满足时，规则5才适用。

请注意，规则1—4采取的是准祈使句的形式，即只有当X时才能说出P这种形式；规则5采取的是把说出P定为Y的形式。因此，规则5是我们在第二节讨论的构成规则系统中特有的一种规则。

还请注意，那个让人听厌的游戏类比显然还可以继续沿用下来。如果我们要求我们自己按一个下棋者的条件去正确地移动马，那么我们会发现诸如必须要轮到自己才能走这样的准备条件，还能发现像陈述马所能移动的实际位置这样的基本规则。我认为在比赛游戏中甚至存在着一种真诚规则，它是双方力争获胜的规则。我认为，比赛的一方在赛场上有意放弃竞争，这种行为很类似于说话编造虚假允诺的情况。当然，在游戏中通常没有命题内容规则，因为总的来说游戏不描绘事态。

如果这种分析适用于允诺情况以外的任何一般情况，那么这些特性应该也存在于其他类型的言语行为中，我认为只要稍加思考就可看到这种情况。例如，请考虑下命令的情况。其准备条件包括说话者应该处在一种听者之上的权威位置上；其真诚条件是说话者希望被命令的行为得到完成；其基本条件必须涉及到这一事实，命令的表述就是企图让听者照它去做。再比如断言，其准备条件包含这一事实：说话者必须具有假设所断言的命题为真的某种根据；其真诚条件是他必须相信这一断言将会为真；其基本条件必须涉及这一事实：断言的表述就是企图告知听者并使他相信他断言的真理性。问候是一种更为简单的言语行为，即使在这里，我们分析过的一些特性也是适用的。在说"你好"时，没有命题内容，也没有真诚条件。其准备条件是说者

必须与听者相遇，其基本规则是作问候表明有礼貌地认出了听者。

如果要对这个问题作进一步研究就要求我们把类似的分析运用到其他类型的言语行为中去。这不仅能使我们对与它们自身有关的概念进行分析，而且对不同的分析加以比较将会加深我们对整个论题的理解。此外，这将为我们得到一种比任何通常信手拈来的分类（如评价的与描述的或认知的与情感的）更为认真的分类提供一种依据。

意向性及其在自然界中的地位[①]

一

意向性是某些心理状态和事件的特征，它是心理状态和事件（在以下这些词的特殊含义上）指向、关于、涉及或表现某些其他客体和事态的特征。例如，如果罗伯特相信罗纳德·里根是总统，那么他的这一信念就是一种有意向的状态。因为从恰当的意义上说，他的这一信念指向、关于、涉及或表现了罗纳德·里根以及罗纳德·里根是总统这一事态。在此情况下，罗纳德·里根是罗伯特这种信念的意向对象，并且罗纳德·里根作为总统这一事态的存在是他的那种信念的满足条件。如果一种信念什么也没有涉及，那么它就没有意向对象；如果它没有获得它所表现的事态，那么这种信念就没有得到满足。

意向性的归属有不同的种类，这些不同正是产生混乱的一个根源，因此我先来对它们作一些区分。请考虑说出以下语句时所作的陈述：

A. 罗伯特相信罗纳德·里根是总统。

B. 比尔看见天正在下雪。

C. "Es regnet"意思是下雨。

---

① 选自《综合》第61卷（1984年），第3—16页。

D．我车子里的恒温器感觉到了发动机温度的变化。

以上每一个陈述都有一个意向性的归属问题，但它们的归属情况有所不同。A完全把一种有意向的心理状态，一种信念，归属于个人；B就不仅仅如此，因为说某人看见某物就意味着他不仅仅具有某种意向性的形式，而且他的这种意向性还是被满足的，也就是说它还实际获得了满足条件。"看见"，与"知道"一样，而不同于"相信"，它是一个成功动词：X看见P蕴含着P。B表明了一种意向现象，因为一个人为了要看见某物就必须有视觉经验，并且这种视觉经验是意向性的载体。但是B不仅仅表明了一种视觉经验，它也表明这种视觉经验是得到了满足的。此外，视觉经验不同于信念，它是有意识的心理事件而不是心理状态。我们完全可以说一个睡着的人相信这样或那样的事，但不能说他看见了这样或那样的事。信念与视觉经验都是存在于行为者心与脑之中内在的意向现象。说它们是内在的仅仅是指这种状态和事件确实存在于行为者的心脑之中。这些状态与事件的归属，属于名副其实的归属，它与讲话方式不同，也与速记一种描述行为者身外更复杂的一系列事件与关系的陈述不同。

再来看后两个语句，A和B中的意向性归属与C和D有所不同。尽管C中的意向性不是内在的，而是派生的，但C确实还应归于意向性。"Es regnet"意思是下雨，这一点儿也没错，但这种意向性不是该语句内在固有的。这个语句可以指别的事情或什么也不指。这种意向性形式属于对某个陈述的速记或者属于讲德语的说话者用这个句子来意指一事物否定另一事物的陈述，并且这个语句的意向性是从讲德语的说话者更基本的内在意向性形式中派生出来的。另一方面，在D中，不存在真实的意向性归属，因为我车子的恒温器确实不具有感觉。D与C不一

样，它属于一种隐喻的意向性归属；但与C相同的是，它的旨意取决于行为者某个内在的意向性。我们用车子里的恒温器调节发动机温度，因此它们必须能够对温度的变化作出反应，于是就出现了这种比喻。如果我们不把对A、B和C的分析与D混淆起来，那么这种比喻是毫无害处的。

总之，仅从这几条陈述就出现了这样一些区分，除了通常对有意识的与无意识的意向性形式的区分以外，我们还需记住：

1. 表明意向现象被满足的那种意向性归属与没有表明这一点的意向性归属之间的区分，如A与B所示。

2. 意向状态与意向事件之间的区分，也如A与B所示。（有时出于简洁的目的，我把它们两者都称为"意向状态"。）

3. 内在的意向性与派生的意向性之间的区分，如意向性在A、B中的归属与在C中的归属双方的区别所示。

4. 像A、B和C这样的真实的意向性归属与像D这样的隐喻的意向性归属之间的区分。前者，不论是内在的还是派生的，其真实性取决于某种意向现象的存在；后者，尽管其隐喻的旨意可能出于行为者某种内在的意向性，但它根本不能归属于真实的意向性。

在本文的以下篇幅里，我只讨论内在的意向性，大致地说，我所要讨论的问题就是：什么是内在的意向性在自然界中的地位？

二

意向心理现象是我们自然生物生命史的一个组成部分。感

觉口渴、具有视觉经验、怀有愿望、恐惧和期待与呼吸、消化、睡眠一样都是一个人生物生命史的组成部分。意向现象与其他生物现象一样，是某些生物机体的真实内在特征，也就是说它们与有丝分裂、减数分裂、胆汁分泌一样是某些生物机体的真实内在特征。

内在意向现象是由在脑中进行的神经生理过程导致的，并且是在脑结构中发生和实现的。关于像突触的神经元放电怎样导致了视觉经验和渴觉这样一些问题的细节，我们所知甚少。但对此我们并不是什么也不知道，就这两种意向现象来说，我们甚至已经找到了说明它们在脑中部位的很好证据。这就是说，至少对于有些意向现象来说，我们已经知道了某些脑器官所起的特殊作用，比如，在产生意向现象时视觉皮质与下丘脑所起的特殊作用。对于我们现在的讨论，更重要的在于意向性在脑中到底是怎样一种过程，关于这一点我们的无知是一种有关细节的经验上的无知，而不会由于"心"与"脑"这两个不可比较的范畴形成一条形而上学的鸿沟妨碍着我们永远去克服这种无知。的确，我们可以从自然界的其他部分中看到我们非常熟悉的那些包括心理现象与脑关系在内的一般关系。在自然界领域中，发现一个系统较高层次的特征是由较低层次的微观客体的行为导致的，并且是在微观客体系统的结构中体现出来的，这是很普通的事。例如，就拿我现在在打字机上正敲打的金属块来说，其固体性是由构成金属的微观粒子行为导致的并且是在微观粒子的系统结构中，即原子与分子的系统结构中体现出来的。固体性是系统的一个特征而不是任何个体粒子的特征。与此相似，就我们所知道的脑而言，心理状态是脑的特征，它是由微观层次上元素的行为导致的，并且是在微观元素即神经元的系统结构中体现出来的。一种心理状态是神经元系统的一个特征，而不

是任何特殊的神经元的特征。而且，按照这种观点，把心理状态视为副现象决不比把世界上任何其他内在的、高层次的特征，比如这台打字机的固体性，视为副现象具有更多的理由。

总之，某些生物体具有内在的意向状态，这些意向状态是由这些生物体神经元系统中的过程导致的，并且它们是体现在这些神经元系统的结构之中的。人们从自然主义的意义上理解下述主张：生物机体消化食物；消化是由消化道中的过程导致的；并且是在胃与消化道的其余部分中完成全部消化过程的。我们的那些主张也应该从这种自然主义的意义上来理解。从自然主义的意义上来理解我们上述观点，一部分困难来自这样的事实：我们用来讨论这些难题的传统术语是围绕着一个17世纪的"心身问题"的概念提出来的。如果我们坚持继续使用传统术语，那么我们可以说：假如二元论是一种属性二元论的话，一元论与二元论就是完全相容的；假如我们认识到心理属性与许多其他种类的属性一样只是一种较高层次的属性的话，那么属性二元论与完整的物理主义就是相容的。这种观点与其说是二元论的，不如说是多元论的，它断言内在心理属性在许多属性中只是一种较高层次的物理属性（这一点也许是完全不去使用那些传统术语的一个很好的理由）。

人们时常否认内在意向现象的存在，这是当代理智生活中的一个显而易见的事实。有时，人们认为具有意向状态的心是某种抽象的东西，比如像计算机程序或流程图表那类的东西；或者认为心理状态不具有内在的心理状况，因为它们完全可以通过它们的原因和结果来加以定义；或者还认为不存在任何内在的心理状态，谈论心理状态不过是我们对付环境的一种讲话方式；有时，人们甚至认为，根本不应该认为心理词汇代表了世界上的实际事物。如果找到了人们持有这些看法的理由，找到了

人们否定生物学向我们提供有关脑的知识的理由，那么也就会找到导致我们当前理智混乱的某些原因。造成这种混乱的原因之一（尽管这只是原因之一）是人们顽固地持有这种信念：如果我们承认了内在意向状态的存在，那么我们将面临着一个不可解决的"心身问题"或"心脑问题"。但是，按照达尔文的意思，脑导致心理现象并不比物体具有引力更神秘。面对这两个事实，人们可能并且确实也应该具有一种畏惧和神秘感，但是这种感觉并不能让我们相信，否认心理状态的存在比否认引力的存在具有更充分的理由。

有些哲学家认为我仅仅断言世界上存在着内在的意向心理状态和事件是不能说明问题的。他们反驳说，因为正像太阳在不动的大地上升落的现象被表明是一种幻觉那样，科学的进步不是也可能表明你的断言将成为一种幻觉吗？相信内在心理状态与相信地球是一平面并在宇宙中有它固定的位置一样不都是前科学的吗？①

但是如果我们暂时把我们的注意力限制在有意识的意向心理事件与状态上（它们毕竟是意向性的初级形式），那么我们就会看到把相信地球是平面的、不动的这一信念与相信心理现象存在这一信念相类比是不能成立的。对于地球来说，事物实际怎样与它们表面看上去怎样有着清楚的区别，但对于有意识的意向现象那种存在来说，就很难找到一种与此相类似的区别。不管怎么说我理解以下说法的含义：虽然地球看上去是平面的、固定不动的，但事实上它不是平面，不是不动的，而是圆的、运动的；但我无论如何也不能理解这样的话 虽然在我看来我现在是

① 见R·罗蒂，"无法言谕的心"一文，载R·Q·埃尔维编《心与自然》一书（哈珀和罗出版公司，旧金山），第84页。

有意识的，但事实上我确实没有意识，而是……什么呢？

我们之所以不能够用看上去不荒谬的东西来接着回答这一问题，其原因自笛卡尔以来就已为人们所熟悉了：如果我有意识地认为我是有意识的，那么我肯定就是有意识的。正是由于这个原因，使我们不可能把这种有意识的心理状态的存在一般地区分为事情"看上去怎样"与"实际上怎样"。

这当然不是说我们不能发现关于我们精神生活、关于有意识与无意识心理状态的性质与机制的各种各样的令人惊奇的和反直觉的东西。这里要指出的是，我们不能把事情看上去怎样与它实际上怎样之间的区分应用到我们自己有意识的心理现象的存在上来。

三

由于人们普遍反对从自然主义意义上将意识和意向性仅仅视作其他属性中较高层次的属性这种观点；由于本文提出的关于意向性在自然界中的地位的观点又与人们时下接受的观点很不合拍，所以我想对这些观点作一些更深入的探究。如果一个人读了自赖尔的《心的概念》(1949年)出版以来这30多年中关于"心身问题"的权威文献，[①] 他就会发现在这个持续着的争论中有一个奇怪的特点，这就是几乎双方所有的参加者都心照不宣地假定有意识的心理事件的那些明确的心理特征不可能像世界上任何其他较高层次的特征那样作为世界的普通物理特征。这

---

① 见N·布洛克编的《哲学心理学文集》，第1卷(哈佛大学出版社，坎布里奇，马萨诸塞州，1980年)；C·V·博斯特编的《心脑同一理论》(麦克米伦，伦敦，1970年)；D·M·罗森塔尔编的《唯物主义与心身问题》(普兰蒂斯-霍尔，恩格尔伍德克利夫斯，新泽西州，1971年)。

种假定常常通过其陈述方式被掩饰起来。当同一性理论家对我们说心理状态仅仅是脑状态时，听起来好像这种论点与我们关于意识与其他意向性形式内在的不可还原的特征的常识假设是完全一致的。听起来这种论点好像是在说，心理过程不过是像消化过程发生、存在于消化道中那样发生、存在于脑中。但是，一般说来，这并不是同一性理论家们所主张的观点。在我们对他们所讲的原文，特别是在对他们答复他们的二元论论敌的那部分文字加以仔细的研究之后，就会发现一般说来同一性理论家们(唯物主义者、物理主义者、功能主义者等等)最终总要否定世界内在心理特征的存在。J·J·C·斯马特以其典型的坦诚，在答复J·T·史蒂文森时明确地表述了他的立场：

"我的答复是我不承认存在任何这样的P属性（即妨碍我们用物理主义体系中的属性去定义"感觉"的那种感觉属性）"(《心脑同一理论》，1970年，第93页)。

为什么斯马特感到必须否定斯蒂文森和我都相信的这个明显的真理呢？我认为这只是因为与笛卡尔以来的传统一样，他认为把有意识的心理现象作为实在也就等于承认了某些神秘现象的存在，承认了在物理科学彼岸存在着某种"理论的悬置物"。另一方面，一些人通过断言他们当下具有一系列有意识的状态这样一些明显的事实来向唯物主义的传统提出挑战。他们似乎认为他们的这种主张属于某种二元论形式，好像断言了关于我们有意识的生活的明显事实，就等于承认了某些本体论范畴的存在与我们大家生活在其中的这个普通物理世界的存在完全不同似的。前一类哲学家认为，他们自己是在保卫科学的进步反对残存的迷信。后一类哲学家认为，他们断言了任何时刻的内省都能展示出来这些明显的事实。但是这两类哲学家都接受了朴素心理主义与朴素物理主义二者是决不相容的这一假定，他

们都接受了对世界的一种纯物理的描述不能涉及任何心理客体这一假定。

这是一些错误的假定。如果从物理实在的实例中排除了心理现象，那么心理现象和物理实在这两个普通的概念就没有任何意义，除非我们采用一方是另一方的否定的方法来定义"物理的"与"心理的"。

那么，我们再进一步追问为什么双方都犯了这个显而易见的错误呢？我认为答案就是他们忠实地继承了至少可回溯到笛卡尔的整个传统，这种传统使人卷入到关于实体、二元论、相互作用、浮现、本体论范畴、意志自由、灵魂不死、道德预定以及其他这类问题的无休止的争论中去。这种传统在很大程度上围绕着用"心理的"与"物理的"来相互命名其余范畴这一假定。但是假设我们能暂时完全忘掉所有这些传统，那么请想象一下这样的情况：用我们所知道的关于生物学、化学和物理学的东西，用从我们自己的经验中知道的关于我们自己的心理状态的东西，单单去考察我们自己与动物的心理状态、意向等等在自然界中的地位。我相信如果我们能忘掉这个传统，那么关于在自然界中这种状态的地位问题就会有一个明显的答案。它们是某些生物化学系统即脑的物理状态，但是这种观点不主张还原和化约。心理状态具有像意识性、意向性、主观性、高兴、苦恼等等所有它的这些令人自豪和令人烦恼的特征，它与我们从来就知道的那个心理状态没什么两样。

为了使我们的观点不至于被误解，我想以最简明的方式把我的观点陈述出来。我采用最朴素的心理主义形式：内在心理状态是真实存在着的，有些是有意识的，有些是无意识的；有些是有意向的，有些是没有意向的。对那些有意识的人来说，他们的确具有他们从表面上看具有的那种心理属性，因为一般说来，

对于这样的属性在事情实际怎样与表面上看上去怎样之间没有什么区别。我采取最朴素的物理主义立场：世界完全是由物理粒子以及它们之间的多种多样的相互关系构成的。就世界上真实的事物而言，只存在物理粒子以及物理粒子的各种排列。我认为，对这两种观点不加任何修改，原封不动地同时接受这两种观点是完全可能的。的确，关于这两种情况，前者仅仅是后者的一种特例。

四

假如有意向的心理状态确实真实地存在着并且不被某种新的定义解释为某种虚幻的或可消去的东西，那么它们在对自然界的一种自然主义的或科学的描述中起着一种什么作用呢？

正像某类机体具有某类心理状态是一个生物学的事实一样，某些心理状态在机体与自然界其他部分的相互作用中、在产生机体行为过程中起着因果作用，这同样也是一个生物学的事实。有时口渴导致机体去饮水、饥饿导致机体去寻找和吞食食物、性欲导致机体去交媾，这不过是一些生物学的事实。尽管这都是在同一生物学层次上发生的事，但人类的情况更为复杂。一个人所具有的关系到他或她的经济利益的那种信念，可以是他或她怎样在政治选举中投票的一个原因，他或她的文学爱好可以是他或她购买和阅读书籍的一个原因，想去某地而不去别的地方的愿望可以是一个人购买机票、驾车或乘公共汽车的一个原因。尽管包含在意向的心理状态中的这个因果关系的事实很明显，但是这些因果关系中的逻辑结构以及这些因果关系中包含着的去说明人类行为逻辑结构的那些必然的内容却极不明显。

含有意向性的说明具有不同于其他物理科学说明的某些逻辑特征。这些逻辑特征的首要一点是意向因果关系在导致某类动物和人类行为中的特殊作用。意向因果关系的本质特征是，意向状态自身作为原因起着产生它自己的满足条件的作用，或者它的满足条件作为原因起着产生它的作用。一方面作为一种表现，它产生它所表现的东西；另一方面被表现的对象或事态作为原因起着产生表现的作用。我们可以通过考虑某些例子来说明这一点。如果我现在极想喝一杯咖啡，并且我为了满足这个愿望去行动，那么其内容是"我喝一杯咖啡"的这个愿望就导致了我喝一杯咖啡这个事态。在这个简单的、典型的意向因果关系中，那个愿望表现了它所导致的那个事态。人们经常讨论的那个"行动的原因"与它们所导致的行动之间的"内在联系"不过是对意向因果关系的这一基本特征的一种反思。由于原因是对它所导致的那个东西的一种表现，作为原因，对于原因的规定已经间接地是一种对结果的规定了。

有时，意向状态确实是它满足条件的一部分，是它意向内容的一部分。也就是说，如果它要被满足，它就必须要作为原因起作用。例如意向，只有当它们表现的行动是由表现这些行动的意向导致时才能得到满足。在这方面，意向不同于愿望：一种愿望即使不导致它的满足条件也能被满足；相反一种意向只有在导致它自己的其他满足条件时才能被满足。例如，如果我想发财并且我发了财，在这里即使我的愿望不是我发财的原因，我的愿望仍然得到了满足。但是如果我打算去赚一百万美元并且我十分意外地得到了一百万美元。在这种情况下，在我有意识或无意识地得到这笔钱时我去赚钱的这种意向没有起着原因的作用，那么尽管我的意向所表现的这种事态实现了，这种意向本身也没有得到满足，这就是说，我的意向从来也没有被贯彻执行。

意向与愿望不同，它们具有存在于它们的意向结构中的意向因果性；只有它们导致它们所表现的那个行动时它们才能被满足。在这个意义上，它们作为原因是涉及自我的。所以说打算去喝一杯咖啡这个在先的意向，在内容上不同于想去喝一杯咖啡的那个愿望。我们通过将这个在先意向满足条件的表现，即"我喝一杯咖啡与导致我喝一杯咖啡的这种在先的意向"，与以上提到的那个愿望的满足条件的表现相对比就能看清楚这一点。

像愿望和意向这些"意志"的情况具有我们所说的"由世界向人心的符合趋向"(这种状态的目标是让世界向着符合愿望或意向的内容发生变化)和"由人心到世界的因果性趋向"(心理状态导致它所表现的那个世界上的事态)。像感知、记忆和信念这些"认知"的情况，就其符合趋向与意向因果性而言，具有相反的情况。它们具有由人心向世界的符合趋向（这种心理状态不是在世界上去创造一种变化，而是去符合某种独立现存的实在）；并且在意向状态产生的过程中，在意向的因果关系起作用时，它们具有从世界到人心的因果性趋向（如果我正确地对世界上的事物进行观察和记忆，那么它们存在的方式必然导致我按照它们的那种存在方式去观察和记忆）。

五

我相信要充分阐述意向性的作用及其在自然界中的地位，需要对意向的因果性作更多的研究，仅仅停留在以往的和本文所达到的研究水平上是远远不够的。为了让读者了解意向因果关系的重要性，我只想讲对意向因果关系进行简短概括的三点内容，目的在于对人与动物行为作出一种因果说明，并且指出这种因果说明与通常自然科学中我们具有的某些规范说明的标准

模型之间不同的方式。

1. 在任何一种因果说明中，说明的命题内容规定一种原因。但是在意向的说明中，被规定的原因自身是具有它自己命题内容的一种意向状态。因此，在一种意向说明中对原因的规范规定不仅仅规定原因的命题内容，而且它必须在对命题内容(至少对命题内容的某部分)的说明中实际重复在这个因果过程中起着原因作用的部分。例如，如果我买一张机票是由于我想去罗马，那么，这种说明就是："我买机票是因为我想去罗马。"在这里我重复了起着愿望作用的那个命题内容："我想去罗马。"当意向的说明在说明中准确地重复起着原因作用的命题内容时，它们或多或少是充分的。对这个特征还有一个更进一步的推论：在这种规范的说明形式中使用的概念，不仅仅描述了原因，而且这些概念自身必然在原因的形成中起着作用。例如，如果我说那个人投票赞成里根是因为他认为里根当选总统会增加他发财和幸福的可能性，那么只有像发财和幸福这样的概念作为原因的一部分来起作用时，它们才能够在说明中被用来规定原因。

在标准的物理科学中找不到与此相似的特征。如果我用重力引力加上其他像摩擦力这样一些作用在物体上的力来说明一个下落物体的加速度，那么我这个说明的命题内容就要涉及到像引力和摩擦力这样一些事件的特征，但这些特征本身并不是命题内容或命题内容的组成部分。

这种观点在对心理说明的本质进行探讨的历史中是一个为人们所熟悉的观点，但在我看来这种观点并没有被恰当地陈述和理解。我相信这个观点正是狄尔泰当年(1962年)指出理解的方法是社会科学的根本方法时所要说的一个观点，同样这个观点也正是温奇当年(1958年)指出说明人行为所用的概念必须是

被说明的行为者也使用的概念时所要说的一个观点。我认为对意向因果关系的分析会使我们在理论上更深刻地理解狄尔泰和温奇所追求的那些观点。①

2．要使对意向因果关系的陈述有效或在原因上作出说明并不需要去陈述一个有关法则。在像物理学这样的学科中，我们假设对一种现象的因果说明不可能是完全充分的，除非它能通过具体的例证来说明某个或某些一般规律。但是对意向因果关系来说，情况一般就不是这样。即使我们相信存在着可以被任何给定的行为事例所例证的并且可用某些或其他一些词汇加以陈述的法则，但我们能够加以陈述甚至相信的这些法则在使用意向的原因对人行为进行因果说明的过程中并不是必不可少的东西。

3．说明的目的形式是从目的、目标、意图、意向方面去说明一种现象或一些相似的现象。如果目的说明确实是科学说明的一个子集，那么目的说明将表明自然界必然实际包含着目的现象。关于意向性及其在自然界中的地位，我所极力主张的观点是既承认自然界包含着目的现象，同时也承认目的说明是说明某类事件的恰当说明形式。的确，说目的性是自然界的一个内在的组成部分，正是认为目的、目标、意图、意向是某些生物机体的内在特性这种观点的一个直接的逻辑结论，因为根据定义，目的、目标、意图、意向现象都是有目的性的。同样，目的说明作为说明某类事件的恰当形式是我对这些现象加以描述的一个直接结论，因为认为这些现象通过意向因果关系的形式导致事件的观点是目的论所特有的内容。

---

① W·狄尔泰，《历史中的意义》（H·P·里克曼编泽，纽约，1962年）；P·温奇，《社会科学的观念》（劳特利奇和基根·保罗，伦敦，1958年）。

所有被我称作是“目的”的状态都具有从世界到人心的符合趋向和从心灵到世界的因果性趋向。在目的说明中，用具体实例来说明这种状态最能起到说明的作用。请考虑一种动物的行为，比如说，一头狮子正穿过灌木丛向前行进。我们可以说狮子的这一行为是正在向另一头野兽——它的猎物潜近。这种潜近行为是由一组意向状态导致的：它感到饥饿，它想要吃掉这头野兽，它抱着抓住、咬死、吃掉这头野兽的目的打算跟踪它。它的意向状态表现了可能的未来事态；只有当这些事态实现时，它的意向状态才能得到满足（从世界到人心的符合趋向）；同时它的行为又是造成这些事态的一种努力（从人心到世界的因果性趋向）。目的是自然界的一部分这种观点也就是这样一种观点：某些机体包含着具有世界向人心的符合趋向的未来指向的意向状态，同时这些状态能够在造成它们的满足条件的过程中起原因的作用。

目的说明的逻辑特征值得强调，因为按照某些观点，一种目的说明要按照一未来事件的发生去说明一事件，这就好比要用吞食猎物去说明潜近行为一样。① 但是在我看来，这种观点是将事情的先后顺序颠倒了。所有有效的目的说明都是采用意向因果关系的说明方式进行说明的，而且不存在神秘的倒退式的意向因果关系。时间 t_1 的潜近行为要用当下与在先的即 t_1 与 t_0 的意向状态来说明，并且这些意向状态的目标在于 t_2 的吞食猎物行为。

17世纪伟大的科学革命对物理学中目的论的否定是一次解放，19世纪伟大的达尔文革命对在物种起源问题上的目的论观

① 对这一观点的说明请见R·B·布雷思韦特的《科学说明》（剑桥大学出版社，剑桥，英国，1953年）。

点的否定也是一次解放。20世纪在关于人的科学领域中一股势不可挡的倾向是通过否定目的论来完成对人的描述。但是，具有讽刺意味的是，以往的这种解放趋势现在正在被抑制住并且在朝着相反的方向发展。为什么呢？因为人类确实具有愿望、目的、意向、意图、目标和计划，并且它们对人类行为的产生起着一种原因的作用，这些不过是关于人类的一个普通的事实。那些完全承认这些事实的人文科学，如经济学，就比那些一开始就试图否定这些事实的学科，如行为主义心理学，取得了大得多的进展。把不包含意向性的系统当作有意向性的系统来研究是拙劣的科学，同样，把有内在意向性的系统作为没有意向性的系统来研究也是拙劣的科学。